

RP SEQUENCE FROM N.A.
RC SRRAIN-W50.
RA ROSS B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.
RA Hocking D., Webb E.;
RT "Porphyromonas gingivalis polypeptides and nucleic acids.";
RT Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF144076; AAD33930.1; -
SQ SPOQUENCE 451 AA; 49825 MW; 3EAC9366FE5AD74 CRC64;

| | |
|---------------------|--------|
| alignment_scores: | |
| Quality: | 663.00 |
| Ratio: | 2.225 |
| Percent Similarity: | 65.929 |
| | |
| alignment_block: | |
| Length: | 452 |
| Gaps: | 7 |
| Percent Identity: | 34.735 |

```

05-09-303-518D-125 x Q9X6S0  ..
Align seg 1/1  to: Q9X6S0  from: 1  to: 451

```

1 ATGATGTAATAATCAAAAAAGGCTTAACTGTCGCCATCGGGGACGACGCGA 50
4 VALLIELVSTHrLysVLSyGLyLeuAlaLeuAsnLeuLysGLyLysProLe 20
51 GCMAACGCTTTACACAGCGCCGGCCATTCCAGCAAGC...GGCTTGCTG 97
20 uprGluwettLeuAlaGluProAlaGlnSerProThrTyrrAlaValAlp 37
98 GCGAAGAATATGCGGATGTCGCCCTCGATGAAGTAAAGTAAGAAGGCGAT 147
37 roAspAspPheGluGluValIleProLysValAlaThrAlaArgProGlyAsp 53
148 GCCGCAAAAAAGGCGCAAGGCGCTTTGAAGCAAAAAAGAAATCCGGCGCT 197
54 LysValaArgAlaGlySerAlaLeuMetHisLysAlaTyrrProGluMe 70
198 GGTGTTACTGCGCGGCTTCAGGCAAAATCGCGCGATTCAGCTGCGGCG 247
70 TlySerPheThrSerProValSerGlyValIleLeuAlaAsnArgGlyA 87
248 AAAGCGGCTACTTTCAGTACAGTCGATTCGCCGTTGAAGCAACGACGAA 297
87 lAlpLysGlyLysValLeuSerIleGluValLysProAspGlyLeuAsnGlu 103
298 ATTCAG...TTTGAACGTATCACACCTCAAGCGGCGGAACATAAGCGG 344
104 TyrrLysPheProValGlyLysProSerAla.....LeuSerAl 117
345 CGAAGAAGTCCGCCGCAACCTGATCCAAATCCGGCTTTGTGACTGCGCTCG 394
117 aGluGlnIleLysGluLeuLeuLeuSerSerGlyMetTrrpGlyPheIleL 134
395 GCACCGCTCGGTTTCAGCAAAATTCCTGCGCGTCATGCCGACCGCTTCGCC 444
134 ySgInAlpArgProTyrrAspIleValAlaIleThrProAspIleAlaProAsp 150
445 ATCTGTCATAGTCGATGAGACACCAATCCGCTGCTGCCGACCTTCGGGT 494
151 lLeTyrrIleThrAlaAsnPheThrAlaProLeuAlaProAspPheAspH 167
495 CATTATCAAGAAGACGCCGAGAGATTTCAACGCGGCTGTGGTAATGA 544
167 eIleValAlaArgGlyGluGluAlaArgAlaLeuGlnThrAlaIleAspAlaLeu 184
545 GCGGTTTGACCAAGCAAAATTCATGCTGTTGTAAGCAGCTGCCGACAGC 594
184 lAlpLysLeuThrIleGlyLysValTyrrAlaGlyLeuLysProGlySerSer 200
595 GTGCCGCTCGAAATAGCTGCGCAACATGAAACAAACATGAATTTGGGGGCC 644
201 LeuGlyLeuHisAsnAlaGluIleValGluValHis.....GlyPr 214
645 GCATCTCGCGGCTTTGAAGTGACGACCAATCAATTTTCATCAGCGCGGTG 694

```

214 OHISPROIALGLYAsnValIglyValIleuIleasnHisThrLysProIleA 231
695 GCCCGAATAAACCCTGTGGACCATCATATTACAGATTAATTACCAT 744
231 snArGlygluThrValTrpThrLeuLysAlaThrAspLeuIleValIle 247
745 GGGCGTTTGGTTCCACAGCGCCCTCGACACCGGACGGCTGATGCCCT 794
248 GLyArPheLeuLeuThrGlyLysAlaAspPheThrArgMetIleAlaMe 264
795 AGGTGGTTCTCAAGTCACAAACCGCGCCCTTGGCGTACCGTTTGGCG 844
264 tThrGlySerAspAlaAlaAlaHisIleGlyValAlaGlyIleMetProGlyC 281
845 CCAAATATCCGAATTAATCTACGGCGGAATGGTT.....GACACGAGC 888
281 yAsnValPheAlaIAserPheProGlyArgLeuThrIleLysGlySerHis 297
889 AACCCGCTGATTTCGGGTCCGATTTGAAGCGCGCATTCACACAGCCG 938
298 GluArGValIleAspGlyAsnValIleuThrGlyLysLeuGlyIuLy 314
939 GCACGATTATTGGAGCGCTACACATACAGATTTCGTTATCGAAGAAG 988
314 sgluProPheLeuSerAlaArgCysAspGlnIleThrValIleProGluG 331
989 GCCGGAGC...AAAGAGCTGTTCGGCTGGGTGGCGCCGACCGGACAA 1035
331 lYAspAspValAspGluLeuPheGlyTrpAlaIleAlaProArgLeuAspGln 347
1036 TACTCCATCACGCGGTACAAACCTCGGCATTTCCG...AAAAACAAC 1083
348 TyrSerMetSerArgAlaIleTyrPheSerTrpLeuGlnGlyLysAsnLysG 364
1083 CTTCAGATTCAACACAGCCGTCAACGCGGCGGACCGCCCATGTGGCCGA 1132
364 uTyrValLeuAspAlaArgIleLysGlyGlyAlaArgIleMetIleMet 381
1133 TTGTACTTACGAGCGGTGATGGCCTTGATATCTGACCCACCTGCTT 1182
381 eAsnGluTyrAspArgValPheProMetAspIleTyrProGluTyrLeu 397
1183 TTGGCGCATTTAATCGTCGGCGACACGACGCGGACGAGCTGGGGTTG 1232
398 LeuLysAlaIleIleAlaPheAspIleAspLysMetGluAspLeuGlyI 414
1233 CTTCGATTTGAGCAGAAAGACCGCTTGTGTGACGCTGTCTGCCCGG 1282
414 eTyrGluValAlaProGluAspPheAlaThrCysGluPheValAspPhe 431
1283 GCAAAATACGAATACGCGCCCTGTGGCGCAAGAGTGGGAACCATTTGAG 1332
431 eTyrIleGluLeuGlnArgIleValArgGluGlyLeuAspMetLeuTyr 447
1333 AAGGAA 1338
448 LysGlu 449
seq_name: sp_bacteriap:Q9WY86
seq_documentation_block:
ID Q9WY86 PRELIMINARY; PRT; 451 AA.
AC Q9WY86;
DT 01-NOV-1999 (TREMBLrel. 12, created)
DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE ELECTRON TRANSPORT COMPLEX PROTEIN, PUTATIVE.
GN TM0244.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
NC NCB1_taxonomy:2336;
[1]

```

SEQUENCE FROM N.A.
 RC STRAIN-MSB8 / DSM 3109;
 MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
 RA Hatt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 genome sequence of *Thermotoga maritima*.";
 RL Nature 393:323-329(1999).
 DR EMBL; AE01708; AAD3335.1. -
 DR TIGR; TM0244; -
 DR InterPro; IPR001450; 4Fe4S-ferredoxin.
 DR InterPro; IPR001949; Complex1_51K.
 DR Pfam; PF00037; fer4; 1.
 DR PROSITE; PS00198; 4Fe4S-FERREDOXIN; 2.
 KW Iron sulfur; Complete proteome; 7B2D7272CFCEB418 CRC64;
 SQ SEQUENCE 451 AA; 49515 MW; 7B2D7272CFCEB418 CRC64;

alignment_scores:
 Quality: 140.00 Length: 454
 Ratio: 0.648 Gaps: 17
 Percent Similarity: 47.577 Percent Identity: 20.264

alignment_block:
 US-09-303-518D-125 x Q9WY86 ..

Align seg 1/1 to: Q9WY86 from: 1 to: 451

```

49 GAGCAGCCGTTTACGACGCGCCGATACCGAAGTCGGTGGTGG 98
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
35 AsplysProIIeGluArgAlaProLeuProGlnLysValPheValPheIle 51
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
99 CGAAGAAATGCGCGTATGCGCCCTCGATGAAAGTCAAGAGAGCGATG 148
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
51 userAmHisAlaGlnAsnProAlaLysProValSerProGlnLysAsp 68
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
149 CCGTCAAAAAGGCCAAGCTGCTGTTTGAACACAAAAGAAATCCGGCGTG 198
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
68 LuValLysThrGlnValAlaIleGlyLysProGlnLysPheIleSerAla 84
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
199 GGTGTACTCGCGCGCTTCAGGAAATGCGCGCATTCACCGTGGCGA 248
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
85 TyrLeuHisSerProValThrGlyArgValLeuGluIle..... 97
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
249 AAGCGCGTACTT.....CAGTCACTCGTGA 274
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
98 LysLysIleLeuHisProIleLeuGlyLysProIleGlnAlaIleValI 114
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
275 TTGCGCGTTGAGGACGACGCAATC.....GAGTTT 306
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
114 LeGluArgThrSerAspAspGlnTrrPValHisIleGluThrGlyAspPhe 130
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
307 GAAGCGTACGACACTGAAGCGCTGGCAAACTTAAGCGCGAAGAGTGGC 356
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
131 GluArgMetSerLysGlnGluIleLeu.....GluIleI 142
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
357 CCGCAACTGATCAATCCGGTTTGTGACCTGGCGTGGCGACCGTCCGT 406
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
142 eLysLysAlaGlyIleValGlyLeuGlyGlyAlaMetPheProThrHisV 159
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
407 TCAGCAAAATTCCTGCGCGTGCATGCGAGCGCTTCGCATTCCTGCAAT 456
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
159 alLysLeuSerProProProGlnLysLysValAspThrLeuIleValAsn 175
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
457 GCGATGAGACCAATCCGCTGGCTGCCAGACCTACGCTCATTAACAAGA 506
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
176 GlyAlaGlnLysGlnProValLeuThrIleAspHisArgLeuMetLeuG 192

```

```

507 AGCGCGGAGATTCAAACGCGCGCTGTGTATTCAGCGCTTGGACG 556
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
192 uArgAlaAspIleLeuGlnGlyIleLeuIleMetLeuLysVal.... 207
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
557 AACGCAAAATTCATTTGTGAAGCA...GCTGGCGGACAGCTGCGTCT 603
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
208 .....LeuGlyValGlnLysAlaValAlaGlyValGlnSerAsnLys 221
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
604 GAAATGCGCGCAACATGCAACACATCAATTCGGCGGCGCGCATCCGTC 653
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
222 MetAspAlaThrHisAsnLeuLysLysValPheLysGly...TyrProVa 237
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
654 CGGTTGAGTGGCAGCGACATTCATTTCATCGACGCGCTGCGCGAATA 703
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
237 LaspValAlaLeuLeuArgThrLysTyr.....ProGlnGlyAlaGln 252
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
704 AA...ACCGTGGAGCAATC..... 720
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
252 ysdInLeuIleTyrAlaIleThrGlyArgMetValProArgGlyLysLeu 268
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
721 .....AATATCAAGATGTAATTAACAT 743
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
269 ProMetAspValGlyValValAlaGlnAsnValGlyThrCysValAlaVa 285
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
744 TGCGCGTTGTTTGCACACAGCGCGTGTGAACAGCGCGCGATTTGGCC 793
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
285 lLysGlnAlaValAlaValAspLysProLeuValGlnArgGlyMetThrV 302
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
794 TAGGTGCTCTCAAGTCAACAAACCGCGCTCTGTGGTACCGTTTGGGT 843
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
302 alSerGlyAspAlaValAlaLysAsnGlnLysAsnLeuIleValArgLys 318
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
844 GCGAAGTATCGCAATTAATCT.....GCGGCGCAATTTGTTGACACAGA 887
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
319 ThrProValLysAspValIleAspTyrCysGlyGlyLysAspLysnTh 335
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
888 CAACGCGGTGATTCGCGTCCGATTTGAACGCG...CGGATTCACAAAG 934
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
335 rGlnArgValIleLeuGlyLysProMetMetGlyIleSerIleThr.... 350
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
935 GCGGCGACGATTAATTTGGAGCGCTACCAATCAATGATTTCCGTTATCGAA 984
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
351 .....AsnLeuAspIleProValMet 357
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
985 GAAGCGCGGACCAAGAGCTTTCGCTGGCTGGCTGGCGCGGACGACAA 1034
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
358 LysGlyThrSer.....PheLeuProLysLys.... 361
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1035 ATATCCATTCACGCGGTACAAACCTCGGCGCATTTCTGAAAACAAATCT 1084
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
362 ...GlyIleThrAla.....PheLeuProLysLys.... 370
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1085 TCAGATTCACACAGCGCTCAACGCGGCGGACCGCGCGATTCGCGGATT 1134
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
371 .....SerArgProGlnLysProCysIleLeuArgCys 380
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1135 GGTACTTACGAGCGCGTATGATCCCTTGATATCCGCCACCTCGCTTTT 1184
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
381 SerGlnCysValGlnValCysProMetAsnLeuGlnProTyrLeuLeu 397
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1185 GCGCGATTAAATCGTGGCGATACCGACAGCGCGGACGATTTGGGTGCT 1234
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
397 rLeuLeuSerThrLysArgLysTyrAspGlnAlaValAlaGlnAsnGlyLeu 414
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1235 TGGATTTGGACGAGAAACACCTCGCTTGGACCTTCGTCGCGCGGCG 1284
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
414 eLaspCysIleLeuLys.....CysGlySerCysThrTyrThrCysProSer 428
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1285 AATATCGAATAC 1296
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
429 LysIleGlnHis 432
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

seq_name: sp_fungi:004051


```

753 .....SerSerThrThrSerSerProSerAlaProValAlaValS 767
172 ACAGCAGCTGGCCCTTTTGGAGGAGCGCTCCTCTACCTTCATCAG 123
767 etherThyr.....ThrSerProSerAlaLeuValValLeu 780
122 GGGCGCATACGGCATATCT.....TCGCCAGCAA 91
781 SerSerThrSerThrSerProThyrAspIleValTyrSerProSerTh 797
90 CCGCAGCTGGTATGCGCGG.....CCGTCGTAACGGCTTCCTCG 47
797 rPheAlaAlaIleSerSerGlyTyrThrProSerProSerAlaSerVal 814
46 GTCGCCCGGAGTGGCAGGTTAGACCTTTTGTATTTA 6
814 IametSerSerThrSerSerSerSerProThyrAspIleVal 827
seq_name: sp_fungi:09C105

```

```

seq_documentation_block:
ID 09C105 PRELIMINARY: PRT: 1236 AA.
AC 09C105:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE GLUCOMYLASE 1 (ALPHA-1,4-GLUCAN GLUCOSIDASE), EXTRACELLULAR
DE STARCH-DEGRADING ENZYME, BY SIMILARITY TO S. CEREVISIAE STAL, CONTAINS
DE CHITINASE FAMILY SIGNATURE.
GN SPAP1E7.04C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxId=4836;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H;
RC Wood V., Rajandream M.A., Barrell B.G., Seeger K., Harris D.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL590605; CAC36921.1;
DR InterPro: IPR001917; AminoTransf-2.
DR InterPro: IPR001579; Chitinase-2.
DR Pfam: PF00192; Chitinase_2; 2.
DR PROSITE, PS00599; AA_TRANSFERS CLASS 2; UNKNOWN 1.
DR SEQUENCE 1236 AA; 123387 MW; 5A2D33A308B7CDD8 CRC64;
SQ

```

```

alignment_scores:
Quality: 132.00 Length: 471
Ratio: 0.569 Gaps: 17
Percent Similarity: 49.257 Percent Identity: 20.807

```

alignment_block:

US-09-303-518D-125/rev x 09C105

Align seg 1/1 to: 09C105 from: 1 to: 1236

```

1331 TCAATGTTTCAGCAGCTTTCGCAACAGCGGCGGCTATTCGATTTGCC 1282
||| ||||| ||||| |||||
505 SerGluValSerSerThrLeu.....LeuS 513
1281 CGGCGAGAGAGCTGCACAAAGCGAGGCTTCTTCTGTCATTC.... 1236
||||| : : : : : ||||| : : : : :
513 rGlySerSerAlaIleProSerThrSerSerThrProSerSerI 530
1235 .....AAGCAACCAATGCTCGCGCTGCGGATATG.....CG 1200
: : : : : : : : : : : : : : : : : : : : :
530 IeIleSerSerProMetThrSerValIleSerSerSerSerIlePro 546
1199 ACGATTAAATCGCG.....AAAGCAGAGGTGGCGAGATATCAAGG 1156
||| : : : : : : : : : : : : : : : : : : :

```

```

547 ThrSerSerSerSerAspPheSerSerSerIleThrIleSerSerI 563
1155 CATCAGCGGCTGTAATACCA.....ATCG 1130
||||| : : : : : |||||
563 yIleSerSerSerIleProSerThrPheSerSerValSerSerIle 580
1129 GCACCATGGCGGCGCGCGGCTGACGGCTGTTGAACTTGAAAGT 1080
: : : : : : : : : : : : : : : : : : : : :
580 euSerSerSerThrSerSerProSerThrSerLeuSerIleSerSer 596
1079 TTTGTTTTCAGAAATGCGCGAGGTTGTACCGTGATGAGTATTGTC 1030
597 .....Se 597
1029 CCGCTCCGCGGCAACCGGAGCGCTTTGCGGCTTCTTGA 980
: : : : : : : : : : : : : : : : : : : : :
597 rSerThrSerSerThrPheSerSerIleSerThrSerSerProSerSerI 614
979 TAACGGAATCTGATTGTGAGCGTCCAAATGATGCGGCCCTTGT 930
||| : : : : : : : : : : : : : : : : : : :
614 IeSerSer.....SerIleSerSerSerThr 623
929 GTATCGCG.....CCGTCATACCGGAAATCGGCGGT 889
: : : : : : : : : : : : : : : : : : : : :
624 IleLeuSerSerProThrProSerThrSerSerLeuMetIleSerSerSe 640
888 GTCGTGTACCAATTCGCGCGCAGTAAATTTGCACTTTCGCACCA 839
||| : : : : : : : : : : : : : : : : : : :
640 rSerIleIleSerGlySerSerSerIleLeuSerSerIleSer.... 655
838 AAAGGTACGCAAGAGCGCGGTTGTTGACTTGAGAACCCATGAGCA 789
||| : : : : : : : : : : : : : : : : : : :
656 ..ThrIle.....ProIleSer 660
788 ATACGCGGCTCGGTGTCAGAGCGCTTTCGCAACAAACGCGCAATGT 739
: : : : : : : : : : : : : : : : : : : : :
661 SerSerLeuSerThrThrSerSerValIleProSerSerThrIle 677
738 AATTACATCTGATTAATGATGCTCCACACGGTTTATTCGCGCGACG 689
: : : : : : : : : : : : : : : : : : : : :
677 ValSerSerSerSerSerLeuIle.....ValSerSerProVal 692
688 GTCGTGAGAAATGAATGCGGCGCAGTCAACCGCAGAGCGCGCG 639
: : : : : : : : : : : : : : : : : : : : :
692 IaSerSerSerSerProIleProSerSerSerSerIleValSerThr 708
638 CCGAATTCATGTGTTTCG.....ATGTGGCAGC 610
: : : : : : : : : : : : : : : : : : : : :
709 TyrSerAlaSerLeuSerAsnIleThrHisSerSerLeuSerLeuThrAl 725
609 ATTTTCAGAGCGCAGCTGCGGCGAGCGCTTACAAACATGATTTGTC 560
: : : : : : : : : : : : : : : : : : : : :
725 aMetSerSerSerSerAlaIleProIleSerValIleSerSerThrLeuI 742
559 GTTCGGTCAACAGGCTCATATACCAACAGCGCGGTTGAATCTCGCG 510
: : : : : : : : : : : : : : : : : : : : :
742 IeThrAlaSerSerSerAsnThrLeuSerSerIleThrSerSerSer 758
509 GCTTCTTG...AATATGACCGTAGGCTGCGCAGCGGATTTGCTGTC 463
||| : : : : : : : : : : : : : : : : : : :
759 AlaIleValSerSerThrThrValSerAsnIleSerSerAsnLeuRose 775
462 CATCGCATTTGACGAG..... 447
: : : : : : : : : : : : : : : : : : : : :
775 rAlaThrIleSerSerGlnSerGlnLeuThrAsnSerSerThrLeuAlaT 792
446 .....ATGGCAGACGCTCGGCAATCGACGCAATTTGCTG 408
: : : : : : : : : : : : : : : : : : : : :
792 hrSerLeuThrLeuSerSerSerSerSerSerSerSerSerSerSer 808
407 AAGGAGCGGTGCGCAGCGAGTCCACAAACCGGATTTGATGAGTTGCG 358
: : : : : : : : : : : : : : : : : : : : :
809 ThrAsnGluTyrAsnThrSerPheHisAlaProThrValSerSerThrTh 825

```

```

357 GCGACCTCTTGGCCGCTTAACTTGGACGCGCTTACGTCGTCAGCTT 308
      ::::::::::::::::::::
825 rserSerSerSerThrThrSerLeuAlaAlaAlaSerGlyVal...Asn 841
307 CAANCTGCAAT.....TCGTCGTCGCTTCA 282
      |||||
841 eRAsSerSerLeuThrSerLeuAlaSerGlyValThrSerValThrSer 857
      |||||
281 ACGCAATCACTGACTGAGTACGCGCTTTCGCGACGATGATCCG 232
      |||||
858 ThrLarThrThrAsp..... 863
231 GCGCATTTTGGCTGAGCGCGCGAGTAAACACGCGCGATCTTTT 182
      ::::::::::::::::::::
864 .SerValThrSerThrThrAlaLeuThrSerGlnGlyProSerSerSerV 860
181 TGTCTTCAACAGCACTGGCTTTTTCAGCGCATCGCTTCCTGACT 132
      ::::::::::::::::::::
880 alValSerSerSerSerSerSerThrThrSerSerSerSerSerSer 896
131 TCTATCGAGGGGCGCATACCGGATCTTCTGCGCAAGCAGCGAGCTTC 82
      ::::::::::::::::::::
897 ValThrSerSerValAlaProAlaValThrSerThrGlySerGlnThrSe 913
81 GGTAAATGGCGCGG 69
913 rSerValValGly 917

```

seq_name: sp_human:060382

```

seq_documentation_block:
ID 060382 PRELIMINARY; PRT: 1791 AA.
AC 060382:
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE KIA0324 (FRAGMENT).
GN KIA0324.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rieke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,
RA Goodwin L., Bryant J., Tesmer J., Melnick L., Longmire J., White S.,
RA Ueng S., Tatum O., Campbell C., Fawcett J., Deaven L.;
RT "Sequencing of Human Chromosome 16p13.3";
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
FT [2]
DR EMBL; AC004493; AAC08453.1; -
FT NON_TER
SQ
SEQUENCE 1791 AA: 191306 MW: 3A7B5530AE95F3E CRC64;

```

alignment_scores:

```

Quality: 131.00 Length: 411
Ratio: 0.697 Gaps: 14
Percent Similarity: 45.742 Percent Identity: 25.061

```

alignment_block:

us-09-303-518d-125 x 060382 ..

Align seg 1/1 to: 060382 from: 1 to: 1791

12 CAAAAAGGTCTAACTGCGCATCGCGGCGACAGCCGAGCAAGCAAGCCGCTT 61

```

776 ArgArgArgSerAlaSerSerProArgThrLysThrSerArgArg 792
      ::::::::::::::::::::
62 ACGAGGCGCC.....GCCATTACCGAAGTCGCGT... 92
      |||
792 YArgSerProSerProLysProArgGlyLeuGlnArgSerArgSerArgS 809
93 .....GTTGGCAAGAAATATCCGCGTATGCGCC 122
809 eArgArgGlyLysThrArgThrThrArgArgArgAspArgSerGlySer 825
123 CTCGATCAAAAGTCAAGAGAGCGATGCCGTCAAAAAGGCGAAGTCTGT 172
      ::::::::::::::::::::
826 SerGlnSerThrSerArgArgArgGlnArgSerArgSerArgSerArgVa 842
173 TTGAGACAAAGAAATCCGGCGGTGCTTTACGCGCGCGCTTCAAGCC 222
      |||
842 LThrArgArgArgArgGlySerGlyThrHis..... 853
223 AAAATCGCGCGATTCACCGTGGCGAAAGCGGACTTCACTAGTCGT 272
      ::::::::::::::::::::
854 .SerArgSerProAlaArgGlnGlnSerSerArgThrSerSer..... 867
273 GATTGCGCTTGAAAGCAAGCAAGCAATGAGTTTAAACGCTACGACCTG 322
868 .....ArgArgArgArgGlyArgSerArgThrProProThr.. 879
323 AAGCGCTGCAAACTTAACGCGCGAAGAGAGTGGCGCGACCTGATCCAA 372
      ::::::::::::::::::::
880 .....SerArgLysArgSerArgSerArgThrSerProAlaProT 893
373 TCGGTTTGGATGCGCTGCGCGACCGCTCCGTTCAGCAAAATCTGTC 422
893 rPysArgSerArgSerArgAlaSerProAlaThrHisArgArgSerArg 909
423 CGTCATGCCGAGCGCTTCGCGATCTGTCGAATGCGATGACACCATC 472
910 SerArgThrProLeuLysSerArgArg..... 919
473 CGCTGGCTGCCGACCTTACGCTATTATCAAGAGCGCGAGAGATTTC 522
920 .....SerArgSerArg..... 923
523 AAGCGCGCTGTTGGTATTGAGCGCTTGCACGCAAGCAAAATCCATGT 572
      |||
924 ..ThrSerProValSer..... 928
573 TTGTAAAGCAGCTGGCGAGAGCTGCGCTGAAATATGTCGCAACATCG 622
929 .....ArgArgArgSer.....ArgSerAr 935
623 AAACGATGATTTGGGCGCGCCGATCTCCGCTTGAAGTGGACGACGAC 672
      ::::::::::::::::::::
935 gThrSerValThrArgArgArgSerArgSerArgAlaSerProValSerA 952
673 ATTCAATTCATCGAGCC.....GGTCGGCGCGCAATA 704
      ::::::::::::::::::::
952 rGArgArgSerArgSerArgThrProProValThrArgArgArgSerArg 968
705 AACCGTGTGACATCAATTATCAAGATGATTAATACCATGCGCTTGT 754
      ::::::::::::::::::::
969 SerArgThrProThrThrArgArgArgSerArgSerArgThrProProVa 985
755 TTGCAAGGCGCGCTGCAACAGCAGCGGTGATTTCCCTAGTGGTGTCT 804
      |||
985 LThrArgArgArgSerArgSerArgThrProProValThrArgArgArgS 1002
805 CAAGTCAACAAACGCGCTTTCCTGACCTTGG..... 841
      |||
1002 eArgSerArgThrSerProLysThrArgArgArgSerArgSerArgThr 1018
842 .....GTGGAAGATATGCAAAATTACTGCGGCGAAATGGTTGACACA 885
      |||

```


[illegible]

seq_name: sp_human:015038

```

seq_documentation_block:
ID: 015038 PRELIMINARY; PRT: 1783 AA.
AC 015038;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA0324 PROTEIN (FRAGMENT).
GN KIAA0324.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBL:Taxid=9606;
[1]
RX SEQUENCE FROM N.A.
RC TISSUE=BRAIN; PubMed=9205841;
RA MEDLINE=97349984; Nakajima D., Ohira M., Seki N., Miyajima N.,
RA Nagase T., Ishikawa K., Nakajima N., Ohara O.,
RA Tanaka A., Kotani H., Nomura N.,
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 4:141-150(1997).
EMBL: AB002322; BAA20782.2; -.
DR NON_TER 1
FT SEQUENCE 1783 AA; 190940 MW; 660302F6FD4179AB CRC64;
SQ

```

```

alignment_scores:
Quality: 131.00 Length: 411
Ratio: 0.697 Gaps: 14
Percent Similarity: 45.742 Percent Identity: 25.061

```

```

alignment_block:
US-09-303-518D-125 x 015038 ..

```

```

Align seg 1/1 to: 015038 from: 1 to: 1783

```

```

12 CAAAGAGTCTAAACCTGCGGCGGAGACCGGAGCAAGCGCTT 61
   ::::::::::::::::::::
775 ArgArgArgSerArgSerArgSerArgSerArgSerArgSer 791
   ::::::::::::::::::::
62 ACGAGCGGCC.....GCCATTACCGAGTCGGCT.... 92
   ::::::::::::::::::::
791 yArgSerProSerProLysProArgLysLeuGlnArgSerArgSer 808
   ::::::::::::::::::::
93 .....GCTTGGCGAAGATATGCGCGTATGCGGCC 122
   ::::::::::::::::::::
808 eArgArgGluLysThrArgThrArgArgArgArgSerArgSer 824
   ::::::::::::::::::::
123 CTCGATGAAGTCAAGAGAGCGGTCGCGCAAAAAAGCGCAGTCTGT 172
   ::::::::::::::::::::
825 SerGlnSerThrSerArgArgGlnArgSerArgSerArgSerArg 841
   ::::::::::::::::::::
173 TTGAAGACAAAAGAAATCCGGCGTGTCTTACTGCGCGGCTTCAGGC 222
   ::::::::::::::::::::
841 lThrArgArgArgArgGlyGlySerGlyTyrHis..... 852
   ::::::::::::::::::::
223 AAAATCGCCGCGATTACCGTGGCGAAAGCGCTACTTCAGTCAGTCGT 272
   ::::::::::::::::::::
853 ..SerArgSerProLysArgGlnLysSerArgSerArgSer..... 866
   ::::::::::::::::::::
273 GATTTCGCTTGAGGCAAGCAAGAAATGAGTTCAGCTACGACCTG 322
   ::::::::::::::::::::
867 .....ArgArgArgArgGlyArgSerArgThrProProth..... 878
   ::::::::::::::::::::
323 AAGCCGTGCAAACTTAAGCGCGAGAGAGTCCCGCAACCTGATCCAA 372
   ::::::::::::::::::::
879 .....SerArgLysArgSerArgSerArgThrSerProLysProt 892
   ::::::::::::::::::::
373 TCCGGTTTGACCTGCGCTGCGACCGCTCCGTTACGCAAAATTCCTGC 422

```

```

seq_name: sp_fungi:094317
seq_documentation_block:
ID 094317 PRELIMINARY; PRT: 534 AA.
AC 094317;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)

```

DT 01-MAY-1999 (Trembl) 10, last sequence update)
 DE 01-MAY-2001 (Trembl) 16, last annotation update)
 DE HYPOTHEICAL 53.4 KDA SERINE-RICH PROTEIN C215.13 IN CHROMOSOME II
 GN SPK215.13.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Lyne M., Kajandream M.A., Barrell B.G., Rieger M.;
 RL Submitted (Nov-1998) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
 CC -1- SIMILARITY: TO YEAST SPAL.
 DR EMBL: AL03534; CAA22127.1.
 KW Hypothetical protein; Glycoprotein; Signal.
 FT SIGNAL 1 14
 FT CHAIN 15 534
 FT CARBOHD 31 31
 FT CARBOHD 426 426
 FT SEQUENCE 534 AA; 53361 MW; D62DC15C80A698E4 CRC64;
 SO

alignment_scores:

Quality: 126.00 Length: 415
 Ratio: 0.603 Gaps: 15
 Percent Similarity: 50.361 Percent Identity: 20.241

alignment_block:

US-09-303-518D-125/rev x 094317 ..

Align seg 1/1 to: 094317 from: 1 to: 534

1280 GGGCAGACGACGATGACCAAGAGGATCTTCGTCACATTCACAGCA 1231
 67 GYGLYTHrProThrTyrTyrGlyTyrAlaThrProThrSerSerG1 83
 1230 ACCCAATGCTCGCGCGCTGCGATGCGCGAGATTAAATCGGCAAAA 1181
 83 uProSerIlePheSerGluSerAlaThrProSerGluThrAsnSerTyrS 100
 1180 GCAGGGTGGGAGATATCCAAAGGCAATCACGCGCTGTAGTATCAATC 1131
 100 erSer.....ProVal 103
 1130 GGCACCATGGCGCGCGCGCTGACGCGCTGTGAATTTGAAGAG 1081
 104 SerSerTyr.....SerAspProAlaThrSerGlnLeuProSerSerTh 118
 1080 TTTGTTTTCAGAAATGGCGGAGGTTGTACGCGGTATGAGATTTGT 1031
 118 rSerPheSerPro.....ThrSerSerGluTyrThrP 130
 1030 CGCGCTGGGCGCAACCCAGCGGACGATCTTTCGCGCGCGCTTCG 981
 130 to.....SerSerThrGluSerSerSerLeuLeuAspProSerSer 143
 980 ATACGGAATCTGATTTGTGTAGCTGCCAAATTAATCGTGGCGCGCTTG 931
 144 ValSerSer..... 146
 930 TGTATGCGCGCGCTTCATACGACGAGAAATCAGCGGTTGTCTGT 881
 147 .AlaIleuProSerSerThrSerValGluVal.....SerIleS 160
 880 CAACCAATTCGCGCGCGATTAATTTGGATATCTTTCGACCAACAAAGGTA 831
 160 erSerSerSer..... 163
 830 CGCAAGAGCGCGGTTGTTGACTGAGAACCAACCAAGGCGATACGCG 781

164LeuSerSerSerAspProLeuThrSerSerThrP 175
 780 CTCGGTGTTCAGACGCGCTTTCGCAAAACGCGCAATGTAATATACAT 731
 175 eSerSerSerSerSerSerThrSerSerSerGlnProSerVal..... 189
 730 CTGTATATATGATGATGATGATGATGATGATGATGATGATGATGAT 681
 190SerSerSerSer 195
 680 AATGATATGTCGCTGCTCAACCGCGAGATGCGCGCGCGCAATTC 631
 196ThrPheSerSerAlaIleProThrSerTh 205
 630 ATGATGTTGATGTTGATGATGATGATGATGATGATGATGATGATGAT 587
 205 rSerSerSerTyrLeuSerSerSerSerSerValValSerSerSerSer 222
 586 CAGCTGCTTACCAACATGATGATGATGATGATGATGATGATGATGAT 537
 222 rSerSerSerSerSerSerThrLeuThrSerSerSerLeuSerThSer 238
 536 AACAGCGCGCTTGAATCTCGCGCGCTTGTGATATGACCGTAG 487
 239 SerIleProSerThrSerSerSerSer..... 248
 486 GTGCGCAGCGCAGGATGATGATGATGATGATGATGATGATGATGATG 437
 249 .SerThrSerSerSerSerSerSerSerSerSerSerThrAlaSerS 265
 436 GCTCGGATTCAGACGCGCAATTTTGTGAAGGAGGCGGTGCGCAGCG 387
 265 erSerSerSerSerSerSerIleSerSerSer.....SerSerSer 279
 386 GTCCAAACCGGATGATGATGATGATGATGATGATGATGATGATGAT 337
 280 SerSerSerPro.....ThrSerThrSerThrIleS 291
 336 GTTTCACGCGCTTACGCGGTGATGATGATGATGATGATGATGATGATG 287
 291 rSerSerSerSerSerSerSerSerProThrSerThrSerThrIleS 308
 286 CTTCAACGCAATCAGCAGCTGATGATGATGATGATGATGATGATGAT 237
 308 erSerSerSerSer.....SerSerPheSerSerThrLeu 321
 236 ATCGCGCGATTTTGCCTGAAGCGCGCAGTAAACCAACGCGCGGANT 187
 322 SerSerSerSerSerSerSerSerSerSerPheSerSerSerPro.... 336
 186 CTTTTCCTTCAACAGCAGCTTGTGCTTTTTCAGCGCAGTCGCTTCT 137
 337ThrSerSerSerThrIleSerSerSerSerSerSerProSers 352
 136 TGACTTTTCATGAGGCGCGATACCGGATATCTTCGCGCAACGAGCG 87
 352 erSerSerPheSerSerThrThrSerSerSerSerSerSerPhe 368
 86 ACTTCGTAATGCGCGCGGTGTAACGCGCTTGTGCGGTCTG 42
 369 SerSerThrValSerSerSerSerSerSerSerSerSerThrLeu 383

seq_name: sp_bacteriap:098DE3

seq_documentation_block:
 ID 098DE3 PRELIMINARY; PRT; 646 AA.
 AC 098DE3;
 DT 01-OCT-2001 (Trembl) 18, Created)
 DT 01-OCT-2001 (Trembl) 18, Last sequence update)
 DE 2-DEHYDRO-3-DEOXYGALACTONATE KINASE.
 GN MRR4742.

OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAF30309;
 EX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 Takeuchi C., Yamada M., Tabata S.;
 RA "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti."
 RT DNA Res. 7:331-338(2000).
 RL EMBL: AP003004; BAB51328.1;
 DR Kinase; Complete proteome.
 KW SEQUENCE 646 AA; 68943 MW; 6AEAF6471BB6FB CRC64;
 SQ

alignment_scores:
 Quality: 120 50 Length: 436
 Ratio: 0.739 Gaps: 23
 Percent Similarity: 37.385 Percent Identity: 23.853

alignment_block:
 us-09-303-518d-125 x Q98DE3 ..

Align seg 1/1 to: Q98DE3 from: 1 to: 646

```

26 ACCTGCCCATCGCGGCAAGCCGAGCAAGCCGTTACGACGCGCGGCC 75
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
23 ThrCysValThrArgMetArgAlaIleHisAlaIleTyrSerAlaAlaArgPr 39
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
76 ATTAACCAAGATCGCGT.....TGCTGGCAAGAAATATCC 110
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
39 oleuGlyArgAlaIleAspAlaValIleTyrIleLeuArgMetIleP 56
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
111 CGGTATCGCGCCCTCGATGAAGTCAAGAGGCGCATGCCGTCAAAAAG 160
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
56 ro.pheArgIysSerIleArgPheSerGlySer..... 66
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
161 GCCAAGTGTCTTTGAAGACAAAAGAAATCCGGCGCTGTACTCGC 210
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
67 .....CysA 68
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
211 CGGCTTCAGGCAAAATCGCGGATTCACCGTGGCGAAAGCGCGTACT 260
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
68 IagIy.....ArgValIleIuIyIasp 74
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
261 TCAGTCAGTGTGATTCGCGTGAAGCAAGCAAGCAAGTGAATGATTGAAC 310
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
75 AlaIleGlyIlePheCysArgProGlyIleArg..... 85
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
311 GCTACGCACTGAAGCGCTGCAAACTTAAAGCGCGCAAGATGCGCGC 360
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
86 .....LeuGlyAlaAspHisArgIleArgPheArgAsp 97
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
361 AACGTGATCAATCGCGTGTGTGACATGCGCGTGGCGACCGTCCGTTAC 410
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
97 IArgGlyAlaAspArgIleVal...CysAla.....ProGlyCysGln 110
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
411 CAATAATTCCTCGCGCTCGATGCGCGCAAGCGTTCGCGCATTCGTAATG 460
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
111 .....CysArgIleuHisArgHisArgIle..... 118
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
461 TGGACACCAATCGCGTGTGTGACATGCGCGTTCGCGCATTCGTAATG 510
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
119 .GlySerGlnHisAlaIleCysArgProHis.....A 129
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
511 GCCGAGGATTCGAACGCGCGCTGTGTGATTCGCGCATTCGCGCAAG 560
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||

```

```

129 rglIuGlyIle.....ArgProAlaAlaLeuProGlnAspArgSer 143
561 .....CAAAATCCATGTTTGA 577
144 AlaArgHisArgGlyAlaAlaArgLeuGlyGlyIleGlyGly..... 158
578 AGCGAGTGGCGCGGACGACGCTGCAAAATGCTGCCAATCGCAACA 627
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
159 .GlyAlaIleArgGlyHisGlyAlaAspGlnGlnCysArgLeuGlyArg 175
628 CATGAATTCGCGCG..... 641
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
175 IArgArgArgArgGlyArgAspGlyValLeuGlyGlnIleProGly 191
642 .....CGCGATCCTCGCGTGTGTGACGCGCGCATTCATTTCA 682
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
192 AspGlnProGlnAlaAlaPheLeuHisArgAlaGlyGlyGlyAlaArg 208
683 TCGACCGCGTGGCGCGCAATTAACCGTGTGACGCGCATTCATTCAGAT 732
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
208 sglIuIagIleArgArg.....ArgLeuAspHisGlnLeu..... 219
733 GTAATTACCATTTGCGCGTGTGTGCAACAGCGCTGCAACACGAGCG 782
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
220 .....HisLeuAspIleLeuProHisGlnProSerArgHis..... 231
783 CGTGAATGCGCTAGGCTGTCTCAAGTCACAAACCGCGCTTGCCTA 832
231 ..... 231
833 CGGTTTGGGTGCGAAATATCGCAAAATTAAGTTCGCGCGGATGTTGAC 882
232 .....AlaGlyLeuHisArgArgGlnIleGlyIle 241
883 ACAGACAAACG.....CGTGAATTCGCGGTGATTTGA 917
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
241 IspArgArgProHisGlnGlyAlaIleArgGlnAlaIleArgGlyProHis 257
918 CGGCGCGATTCACAAAGCGCGCGCATTA.....TT 949
258 ProArgGln...ArgArgArgAlaIleArgLeuGlyAspHisArgIleAla 273
950 TGGGAGCGTACCAATCGATTCGCGTATGCAAGAGCGCGCAAA 999
273 agIlyIleLeu.....GlyAspArgArgSerIleSerProCysArg 288
1000 GAGCTGTTCGCGTGGTTCGCGCGCGCAAGCGCAAAATTAATCATCGCG 1049
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
288 ImlaValHisGlnGlyGlyHisAlaProGly..... 298
1050 TACAAACCTCGCGCATTTCTCGAAAAACAACGCTTCAGTTCAACAGAG 1099
298 ..... 298
1100 CGGTCAACGCGCGGACGCGCGCATGTCGCGCATTCGTTACGAGCG 1149
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
299 .....ArgHisGlyArgHisGly.....AlaVal 306
1150 GTGATGCCCTTGATATCCGCGCGCGCGCGCTTTCGCGCATTTAACT 1199
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
306 alProGlyLeuGlyCysLeuAlaHisAlaAspArgAlaAspArg 322
1200 CGGCGA 1205
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
323 ArgArg 324

```

seq_name: sp_human:Q9H9M1

seq_documentation_block:

ID Q9H9M1 PRELIMINARY:

PRT: 639 AA.

AC Q9H9M1:

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

31 GCAGCTT 24
111
530 roglyphe 532

seq_name: sp_invertebrate:Q9VXM5

seq_documentation_block:

ID Q9VXM5 PRELIMINARY; PRT: 2406 AA.

DT 01-MAY-2000 (TRMBLrel. 13, Created)

DT 01-MAY-2000 (TRMBLrel. 13, Last sequence update)

DT 01-JUN-2000 (TRMBLrel. 14, Last annotation update)

DE CG9056 PROTEIN.

OS Drosophila melanogaster (Fruit fly)

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RA MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikos G.L.G.,

RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,

RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,

RA Burris K.C., Busan D.A., Butler H., Cadeu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferris C., Fertler S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Haris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., Mocherson D.,

RA Merklow G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Strong R., Sun E.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Wang A.,

RA Svyrkas R., Tector C., Turner R., Venter E., Wang A., Wang X.,

RA Wang Z.-T., Wasserman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

DR EMBL: AE003500; AAF4853.1; -

DR FlyBase: FBgn0030702; CG9056.

SO SEQUENCE 2406 AA; 248803 MW; 740EAD78D64F76C CRC64;

alignment_scores:

Quality: 119.50

Ratio: 0.511

Percent Similarity: 47.561

Percent Identity: 21.951

alignment_block:

US-09-303-518d-125/rev x Q9VXM5

Align seg 1/1 to: Q9VXM5 from: 1 to: 2406

1340 CCTTCCTTCATGATGTTTCAGACACTTTCGACACAGCGCGCATTC 1291
1278 ProValLeuProValAlaThnProAsnLeuSerAsnLeuProThrAl 1294
1290 GTATTGCCCGGACAGACAGAGCTGCACAAAGGAGGCTTCGTCGA 1241
1294 nhlsArgSerSerAspSerArg..AsnSerArgGlnSerProAlSerL 1310
1240 ATTCACACACCCCAATCCCTGCGCGCTGCGATATG...CCGACG 1197
1310 euLysSerThnProSerAsnLleGlyLeuAsnValSerMetAlaProTh 1326
1196 ATTAATATG...CGCAAAAGCAGGCTGGCGAGATATCCAAAGGCAT 1153
1327 LeuArgSerLleThnProLeuAsnAsnSerSerAlaLleSerSerGlyAl 1343
1152 CACGGCTGCTGTAAGTACCAATCGGCACCATGCGCGCGCGCTTGA 1103
1343 aserGln...ProValSerValValProSerAlaAsnSerL 1357
1102 CGGCTGCTGTTGAATTAAGAGCTTGTTCAGCAAAATGCGCGAGCTT 1053
1357 hAla...LeuSerMetSerAsn...ProhIsLle 1366
1052 GTA...CGCGTATGAGTATTTGTCGCGC... 1026
1367 SerhIsSerhIsHisValProAlaThnAlaSerGlyAlaPhnSerSer 1383
1025 ...TCGGCGGACCCAGCCGACAGCTCTTGT...CTGCGCG 989
1383 rAlaAlaAlaGlyThnSerThnProAsnSerGlyLeuSerThnLeuAlav 1400
988 CTTCCTTCGATACGGAATCTGATGTTGCTAGCTCCCAATAATCTGTC 939
1400 alPhnSerLeuSerThn...SerAla 1407
938 GCGCGCTGCTGTAATCCGCGCTCAATACCGACGGAATACGCGGCTT 889
1408 AlAlPro...GlnProhIsSerhIsPhnProGlnSerThnGlnMe 1421
888 G...TCGTGTG... 882
1421 tLeuProGlnSerGlyAsnPhnSerSerValSerhIsLeuThnThn 1438
881 ...TCACCAATTCGCGCGCAGTAAATTCGATCTTCGCA 843
1438 tSprometSerSerGlnAsnGlnProMetValArgGlySer... 1452
842 CCCAAACGCTACCAAGACGCGCGCTTTCAGACGCGCTTTCGAAACGCTT 793
1453 ...ThnLeuYrSerGlnSerSerAlaAlaAlaThnAlaProhIs 1467
792 GGCATACACGCGCTGCTTCAGACGCGCTTTCGAAACGCGCA 743
1457 rAlaAlaAla...AlaValSerAsnPhnThnProhIs 1478
742 TGATATATACATCTGTAATATGATGCTCCACACGCTTATTCGCGCG 693
1478 ervaLleu... 1480
692 ACCGCTGATGAATGAATGCTGCTGCACCTCAACCGCAGAGATCGG 643
1480 ... 1480
642 GCGCGGATTCATGATGTTTCGATGTCGACATTTTCAGACGCGCAT 593
1481 ...AlaValGlnSerLeuThnThnAlaValThnSerSerSer 1494
592 CTGCGCGACGCTGCTTACAAACATGATTTGCTGCTGCTC... 552

```

1494 erSerProSerThrIleuSerSerSerValIleGlnIlyValIleSerPro 1510
|||||.....:|||||.....:|||||.....:|||||
1511 LysSclgluSerProCysAnlysAspArgspSerSerTySerSerPr 1527
||| ..... ||| :||| |||||
1510 GCGCTTTTATATGACGGTAGGGGTGGCCAGCCAGCGATTGGTGCA 461
||||: ..:::||||:.....: ..::: ..::: |||||
1527 oAlasInlaIvalThrThrCysAlaProThrThrProIleValSer. 1543
oAlasInlaIvalThrThrCysAlaProThrThrProIleValSer. 1543
460 TGCATTGACGAAGATGGCAAAGCGCTGGCATGACGACAGAATTGG 411
::::||| ||| ::|||
1544 ..... SerGlySerAlaIaIproThrProProIleu 1553
::||| |||
410 CTGNAC ..... GGAGGGTGGCGACGCCACTCCACA 379
||| ||| |||:||||| ||| ||| |||
1554 SerAnscYshSerMetelyIleIyMetValasnaIala.SerThra 1570
378 ACCGATTGGATCAGGTTGGCGGCACCTTCCTTCGCCCTTAAGTTGCCA 329
||| |||:||||:.....: ||| ||| |||:.....:
1570 IaArg..... SerSerCysAnaIaIe.. SerProIeuSerIlePro 1584
328 GCCTTCAGT ..... GCGTAGCGTTCAAC 303
:||||:||||| .....: |||:.....:
1584 lahrIaIaIylIehIsvaIserAlaThrAnProSerPheGlnSer 1600
302 TCGATTTCGTGTCCT ..... 285
||| |||
1601 SerTyPheProThrProIeuAlaProProSerSerProSerProAl 1617
284 ..... TCACGGCAATCACAGACTGCATAAGTACGCGCTTTGGCCA. 243
||||| ||| |||:.....: |||:.....: |||:.....:
1617 aThrSerSerAlaIaIleIleSerSerSerAlaSerGlnPheAnPro 1634
242 ..... CGGTATCGCGGCACTTTGGCTCAACGGCGGCACTAAC 201
|||: ..:::||||:.....: ||| ||| |||:.....:
1634 laValSerHisSerMetSerSerIleValIthrAlaIyAlaThrThr 1650
200 ACCAGCCCGGATTCCTTTTGTCTTCAACAGCAGCATGGCTTTTGGAC 151
||||| ||| |||:.....: |||
1651 ThrThr..... AlaSerSerValThr..... 1657
150 GGCATCGCTTCCTTGACTTCATCGAGGGCGCATACCGGCAATVTCTT 101
:::||||:.....: |||:.....: ..:::
1658 .... GlnProSerValaIaIaIleSerAnProValIThrAnThrProH 1673
100 CCCAGCAACGCGACTTCGGA 78
||| :||| |||:.....:
1673 IsProPheSerAlaGluSerLeu 1680

seq_name: sp.Plant:Q9SUVO

seq_documentation_block:
ID Q9SUVO PRELIMINARY; PRT; 857 AA.
AC Q9SUVO;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DR HYPOHERETICAL_96.9 KDA PROTEIN.
AN FB84.120 OR AT4G32420.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
ON NCBI_Taxid-3702;
RX [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseyille R.,
RA De Clerck R., De Keyser A., Neyt P., Roupe P., Van Den Daele H.,
RA Viallard R., Gielens J., Van Montagu M., Hohnel J., Meves H.W.,
RA Mayer R.F.X., Lemcke K., Schellier C.
Submitted (DEC-1998) to the EMBL/Genbank/DBD databases.
```

RN [2] EU Arabidopsis sequencing project;
 RP Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Terryn N., Ardiles W., Buysshert C., Dasseville R., De Clerck R.
 RA De Keiser A., Neyt P., Rouze P., Van Den Daele H., Villacel R.,
 RA Gielen J., Van Montagu M., Mewes H.W., Lemcke K., Mayer K.F.X.,
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL034567; CAA22566.1; -
 DR EMBL; AL161581; CAB79959.1; -
 DR HSSP; P05092; 2CPL.
 DR InterPro; IPR002130; CSA_Piase.
 DR Pfam; PF00160; pro_lisomerase; 2.
 DR PRINTS; PR00135; CSAPIISMASE.
 DR PROSITE; PS50072; CSA_Piase_2; 1.
 RW Hypothetical protein.
 SQ SEQUENCE 857 AA; 96854 MW; 17EAPa842D58D972 CR664;

```
alignment_scores:      Length:    393  
                        Quality:     118.00  
                        Ratio:         0.663  
Percent Similarity:   45.293  
                       Percent Identity: 25.700
```

```
alignment_block:
  ne-00-202-1102
```

US-09-303-518D-125 x Q9SUVc

Align seg 1/1 to: Q9SUV0 from: 1 to: 857

```

24  AAACCTCCCGACCGGGGACGACGAGCAGCAGCCGCTTACGACGCGCCG 73
    : : : : : : : : : : : : : : : : : : : : : : : : : :
520 ArgSerAlaSerArgIysLeuGlyArgGlyProLeuArgIysSer 536
    : : : : : : : : : : : : : : : : : : : : : : : : : :
74  CCATTACCGAAGTCGCGTTCTGTGGCGAAGATATATCCGATACGCCGCC 123
    : : : : : : : : : : : : : : : : : : : : : : : : : :
536 T.....ArgArgSerProSerArgSerPro 545
    : : : : : : : : : : : : : : : : : : : : : : : : : :
124 TCGATGAAGTCACAGGAAGCGCATGCCGCAAAAAGGCCCAAGTCGT 173
    : : : : : : : : : : : : : : : : : : : : : : : : : :
545 AlaArgSerSer.....ArgArgSerLeuSerArgSerProIleGlnLeu 559
    : : : : : : : : : : : : : : : : : : : : : : : : : :
174 TGAAGACAAAAGATCCGCGCGGTGTGTTACTGCGCGCGCTTCAGCA 223
    : : : : : : : : : : : : : : : : : : : : : : : : : :
560 .....SerArgArgSerLeuSerArgSerProThrArgLe 571
    : : : : : : : : : : : : : : : : : : : : : : : : : :
224 AAATCCGCGCGGAT.....TCACGCTGGCGAAGAACCGCTACTTCAG 264
    : : : : : : : : : : : : : : : : : : : : : : : : : :
571 userArgArgSerLeuSerArgSerProIleArgSerProArgIysSer 588
    : : : : : : : : : : : : : : : : : : : : : : : : : :
265 TCAGTCGTGATTTGCGGTGTAAGCAGCAGCAGCAAAATTCAGTTGAAGCTA 314
    : : : : : : : : : : : : : : : : : : : : : : : : : :
588 alSerArgSer..ProValArgSerSerArgIysSer..... 599
    : : : : : : : : : : : : : : : : : : : : : : : : : :
315 CGCACTGAAAGCGGTGCGCAACTTAAGCGGCGGAAGAAAGTCGCCGAC 364
    : : : : : : : : : : : : : : : : : : : : : : : : : :
599 ..... 599
    : : : : : : : : : : : : : : : : : : : : : : : : : :
365 TGAATCCAAATCCGGTTTGTGACTCGCTGCGAGCAGCCGTCGTTACCA... 412
    : : : : : : : : : : : : : : : : : : : : : : : : : :
600 .....ValSerArgSerProValArgSerSerArg 609
    : : : : : : : : : : : : : : : : : : : : : : : : : :
413 ..AAATTCCTGCGCGTGAATGCCAGACCGTTGCCATCTTCGTCATCGCAT 461
    : : : : : : : : : : : : : : : : : : : : : : : : : :
609 gAtArgIleSerArgSerProValArgSerSerArgIysSerValSerA 626
    : : : : : : : : : : : : : : : : : : : : : : : : : :
462 GGACACCAATCCGCTGCGTGGCTGCCGACCTTAGGTCATTATCAAGAACGC 511
    : : : : : : : : : : : : : : : : : : : : : : : : : :

```



```

626 rgSerProIleArgLeu.....SerArgArgSerIleSerArgSerPro 640
512 CGAGGATTTCAACGGCGCTGTGGATTGAGCG.....548
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
641 lIeArgLeuSerArgArgSerIleSerArgSerProIleArgLeuArg 657
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
549 .....TTTGACCGCAAGCAAAATTCATGTTGTAAAGCAAGCTGGCCAGA 593
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
657 rgArgIleSerArgSerProIleArgLeu.....ArgArg 668
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
594 CGTGGCTGTGAATGCTGCAACATGAAACATGATTCGGCGGCC 643
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
669 ArgSerValArgPro.....ArgSerPr 676
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
644 CGCATCTCCGCTGTGAGTGGACGACCATTCATTTCAGAGCGGTC 693
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
676 oProProAspArgArgArgSerIleSerArgSerProIleArgLeu 693
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
694 GCGCGCAATRAAACCGTGTGACCATTCATTTCAGATGTAATTCAT 743
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
693 rgIleArgArgGlyArgGlyPheSerGln.....ArgPheSerTy 706
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
744 TGCCCGTTGTTGCAACAGCGCGCTGTGAACACGAGCGCGTGAATGGCC 793
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
707 .....AlaArgArgTyArgThrSerPr 714
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
794 TAGGTGTTCTCAAGTCAACAAACCGCGCTGTGCGTACCGTTGGGT 843
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
714 O.....SerProAspArgSerPro.....TyArgPheSer 724
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
844 GCGAAATATCGCAATTAAGTGGCGGCGAATGTTGACACAGACAGACCG 893
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
725 .....AspArgSerasp 728
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
894 CGTATTCGCGTTCGATTATGACGCGCGGATTCACAGAGCGCGACG 943
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
729 ArgAsp...ArgPheArgSerArgArgArgPheSerProSerArgPheArg 744
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
944 A.....TTATTTGGGACGCTACACATCAGATTCCTGTATCGAAGA 987
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
744 gSerProIleArgGlyArgThrProIleSerMetLeuArgArgGly 761
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
988 GCGCGAGCAAGAGAGTGTGGCTGGCTGGCGCGACCGGACAAATA 1037
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
761 rg.....SerArgSerValSerProIleArgGlyArgGlnArg 774
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1038 CTCATCGAGCGTACACCT.....CGGCAATTTCCTGA 1072
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
775 ...ArgTySerArgSerProIleArgSerArgSerProIleArgGly 790
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1073 AAACCAACTCTTCAGTTCACACAGCGCTCAGCGCGCGCGCGCC 1122
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
790 sArgArgSerProSerAlaSerHisSerLeuSerProSerArgSerArg 807
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1123 ATGTGGCGATGTGTTACTACGAGC 1148
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
807 erArgSerLySerTySerLySer 815
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: sp_bacteria:093LB1
seq_documentation block:
ID 093LB1 PRELIMINARY; PRT: 299 AA.
AC 093LB1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE PUTATIVE DNA-BINDING PROTEIN.
OS Bifidobacterium animalis (Bifidobacterium lactis).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=28025;
RN 11
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-ATCC 27536;
RA Gonzalez Vana A., Rossi M., Altomare L., Gottl R., Matteuzzi D.;
RT "Effect of pH and culture redox potential on the continuous culture of
RT Bifidobacterium animalis ATCC 27536, studies on stability of
RT recombinant plasmids and characterization of the insertion sequence
RT IS1999."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, AJ18089; CAC41624.1; -.
KW DNA-binding.
SQ SEQUENCE 299 AA; 34340 MW; B71D969CEC0A7A03 CRC64;

alignment_scores:
  Quality: 115.50      Length: 385
  Ratio: 0.726        Gaps: 21
  Percent Similarity: 41.299      Percent Identity: 23.896

alignment_block:
US-09-303-518D-125 x 093LB1 ..

Align seg 1/1 to: 093LB1 from: 1 to: 299

81 CGAGTCCGCTGCTTGGCGAAGATATCGCGGATGCGCCCTGATGA 130
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32 ArgAlaArgValAlaProArgArgValArgArg.....ProVal..... 44
131 AAGTCAGAGAGCGATGCGCTCAAAAAAGCGCAAGT...GCTGTTGAA 177
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
45 .....ArgAspArgAspArgArgProAspProAlaLeuArg 57
178 GACAA.....AAGATCCGCGCGCTGTGTTACTGGCGCGGCTGAG 221
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
57 rgGlnAlaGlyThrAlaGlyHisGlyAla..... 67
222 CAATAATCGCGCGATTCACCGTGGCGAAGAGCGGCTACTGAGTCAGTCG 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
68 .....GlyProPheArgProGlyAlaSerArgMetAspAl 79
272 TGATTCGCTGTAAGGCAACGAGCAATCGATTTGAACGCTAGCACCT 321
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 agAlaAlaArgArgArgValHis..... 86
322 GAAGCGCTGGCAAACTTAAGCGGCGAAGAGTGGCGCGCAACCTGATCA 371
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 .....LysThrArgArgGlyGly..... 92
372 ATCCGTTGTGGATGCGCTGGCGACCGCTCCGTTGACAAATTCCTG 421
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
93 .....ArgHisGlyArgVa 97
422 CGTGCATGCGGAGCGGTTGCGCATCTTGCATTCGATGAGACCAAT 471
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
97 lArgArgLeuGlnAlaGlyArgGlyIleArgArgAlaPro..... 110
472 CGGCTGGCTGCGAGCCCTACGTCATTCATCAAGAGCGCGGAGATTT 521
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
111 .....ArgGlyArgGlyHisArgProValProArgArgAlaThr 123
522 CAAGCGCGCTGTGTGATTGAGCGGTTGACCGCAACCAATTCATG 571
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124 GlyArgArgPro.....ArgHisGlnGlyAlaLe 133
572 TTGTAAGGAGCGTGGCGAGAGTGGCGCTGTGMAATGCTGCCAATC 621
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 uProProAlaThrArgArgHisArgGlnAla.....ArgCysGlnGly 148
622 GAAGACATGATTCG.....CGGCGCGCATTCGCGGCTTGGAG 662
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148 rgGlnAla.....LeuArgLeuGlnAlaArgProAlaAspGlnArg 161
663 TGGACAGCGACATTCATTTCATCGAGCGCGTGGCGGAATAAACCGTGT 712
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
162 .....ArgTl 163

```



```

753 CAAGGCGCAATGTAATTACATCTGATATGATGATGCGACGCGTT 704
    : : : : :
256 rArgHrPro.....ProA 261
703 TATTCGCGCGCGACCGCTCGATGAATGATGCGCGACATCAACG 654
    : : : : :
261 rGAlaserProthrGlySerProPro...ArgAlaserPrometHrPro 276
653 GCAGGATGCGGG.....CC 640
277 ProArGAlaserProArGThrProProArGAlaserProThrHrPr 293
639 GCCGATTCATGTTTCATGTTGCGACGATTTTCAGACGCGCGCTG 590
    : : : : :
293 oSerArGAlaserLeuThrArGThrProSerTrpAlaserProThrHr 310
589 CGCCAGCTGCTTACAAACATGATTTTCGTCGTCGCAACGCGTCAT 540
    : : : : :
310 hrProSerArGAlaserMetLysMetGlnSerThrValSerLeuThr 326
539 ACCACAGCGCGCGTTGAAA..... 519
327 ArGThrProProArGAlaserProThrGlyThrProSerArGAlaserPr 343
519 ..... 519
343 oThrGlyThrProSerArGAlaserLeuThrGlySerProSerArGAlas 360
518 .....TCCTGCGCGCTTTTGTATGATGACGCGTAGGCTG 483
360 erLeuThrGlyThrProSerArGAlaserLeuGlyThrProSerArG 376
482 GCACCGCAGCGANTGCTGCCATCGATTGACGACAGATGCGACGCGTC 433
    : : : : :
377 Alaser.....LeuIleGlyThrProSerArGAlaserLeuThrGly 391
432 GGCATCGACGCGAGATTTGCTGAACGACGCGTGCACGCGCGACGTC 383
    : : : : :
391 rProProArGAlaserLeuThrGlyThrSerSerThrAlaserLeuThrA 408
382 ACAACCGAGATGATGATGATGCGCGCGCATTTCTTCCGCGTTAAGTT 333
    : : : : :
408 rGThrProSerArGAlaserLeuThrArGThrGlnSer..... 420
332 GCCAGCGCTTACAGTGGCTAGCGTTCAACGATTTGCTGCTTC 283
    : : : : :
421 .....SerSerSerLeuThrArGThrProse 429
282 AACGCGATTCAGACTGACTGATGACGCGCTTTTCCGCGAGTGAATCG 233
    : : : : :
429 rMetAlaserLeu.....ThrArGThrProProArg.....A 440
232 CGGCGATTTTG.....CCTGAACGCGCGAGTGAACACGCGCGCC 192
    : : : : :
440 laseLeuThrArGThrProProArGAlaserLeuThrArGThrProPro 456
191 GGATTCCTTTTGTCTTCAACAGACACTGGCCTTTTTCAGCGCATGCC 142
    : : : : :
457 ArGAlaserLeuThrArGThrProProArGAlaserLeuThrArGThrPr 473
141 TTCCTTGACTTCATCAGAGGCGCGACATCGCGCATTTCTTCG..... 99
    : : : : :
473 oSerMetValSerLeuLys...ArgSerProSerArGAlaserLeuThrA 489
98 .....CAAAGCAGCGGACT.....TCGGTA 78
    : : : : :
489 rGThrProSerArGAlaserLeuThrMetThrProSerArGAlaserLeu 505
77 ATGGCGCGGCGCTGTAACGCGCTTGCCTCGGCTCGCC 39
    : : : : :
506 ThrArGThrProSerThrAlaserLeuThrGlyThrPro 518
seq_name: sp_rodent:090Y35

```

```

seq_documentation_block:
ID 090Y35 PRELIMINARY; PRT; 1979 AA.
AC 090Y35;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE HOST CELL FACTOR C1 (FRAGMENT).
GN ABCD1 OR HCFC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Platzer M., Brenner V., Reichwald K., Wiehe T., Okesche A.,
RA Rosenthal A.;
RT "Comparative sequence analysis of the mouse Licam locus and the
RT corresponding region of human Xq28."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF133093; AAF22156.1; -
DR MGD; MGI:1349215; Abcd1.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR01798; Kelch.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF01344; Kelch; 3.
DR SMART; SMO0060; FN3; 1.
DR NON_TER
FT
SQ
SEQUENCE 1979 AA; 203436 MW; ABE98C5DDE68BA9 CRC64;

```

```

alignment_scores:
Quality: 114.50 Length: 618
Ratio: 0.442 Gaps: 25
Percent Similarity: 41.909 Percent Identity: 20.065

```

```

alignment_block:
US-09-303-518d-125/rev x 090Y35 ..

```

```

Align seg 1/1 to: 090Y35 from: 1 to: 1979

```

```

1343 CAGCTCTCTCTTCATGATGTTCCACGACTTTGCCACACGCGCGCTA 1294
    : : : : :
681 LysProThrIleLeuGlyIleSerSerVal.....SerProse 693
1293 TTCGTATTTCCCGCGGACGACGAGCTGCACAAACGAGCTTCTTCTG 1244
    : : : : :
693 rThrThrLysProGlyThrThrThrIleLeuLysThrIlePrometSera 710
1243 CCAATTCACGCAACCCAAATGCTGCGCGCTGCGGTATCGCGAGAT 1194
    : : : : :
710 lalleleThrGlnAlaGlyAlaThrGlyAlaThrSerSerProGlyIle 726
1193 AATTCG..... 1188
    : : : : :
727 LysSerProIleThrIleIleThrThrLysValMetThrSerGlyThrG1 743
1187 .....CGCAACAGCGAGTGGCGAGCA 1166
    : : : : :
743 yAlaProAlaLysIleIleThrAlaValProLysIleAlaThrGlyHisG 760
1165 TATCCAAAGGCGATCAGCGC..... 1146
    : : : : :
760 lylngInglYalThrGlnValValLeuLysGlyAlaProGlyGlnPro 776
1145 .....TCGTAACTACCAATCGGACCAATGGCGGCGGTCGCCGC 1108
    : : : : :
777 GlyThrIleLeuArGThrValPrometGlyGlyValArGLeuValThrPr 793
1107 GTTGACG.....GCTGTTGCACTTCAACAGATTTGT 1076
    : : : : :
793 oValThrValSerAlaValLysProAlaValAlaThrThrLeu..... 806

```

```

1075 TTTTGGAGAAATGGCCGAGGTTGTACGC..... 1047
      |||||:
807 .....ValVallysglyThrThrGlyValThr 816
1046 .....GTGATGAGTATTGTCCGCGTCCGCGC 1018
      |||||:
817 LeuGlyThrValThrGlyThrValSerThrSerLeuAlaGlyAla 833
1017 AACCGAGCCGAACAGCTTTGTGGCGGCTTCTGCATAC..GGAATC 969
      |||||:
833 aHisSerThrSerAlaSerLeuAlaThrProIleThrLeuGlyThrI 850
968 TGATGTGGTAGCGTCCCAATATC..... 943
      |||||:
850 leaIatThrLeuSerSerGlnValIleAsnProThrAlaIleThrValSer 866
943 ..... 943
867 AlaAlaGlnThrThrLeuThrAlaAlaGlyLeuThrThrProThrI 883
943 ..... 943
883 eThrMetGlnProValSerGlnProThrGlnValThrLeuIleThrAla 900
942 .....GTGCGCGGCTTG.. 931
900 roserGlyValGluAlaGlnProValHisAspLeuProValSerIleLeu 916
930 TGTAATCGCGCGTCAATACGAGC..... 904
      |||||:
917 AlaSerProThrThrGlnGlnProThrAlaThrValThrIleAlaAspSe 933
903 .....GGAATACGCGGTTGTCTGTGCACAC 876
933 rGlyGlnGlyAspValGlnProGlyThrValThrLeuValCysSerAsn 950
875 AATTCGCCGAGTATTGGCATCTTCCGACCCCAACGCTACGCA 826
950 roProCysGlnThrHis.....GluThrGlyThrThr 960
825 GAGCGCGGTTGTGACTTGAACAACACCTAGGCAATAC.....GC 782
961 AsnThrAlaThrThrThrValAlaHisLeuGlyGlnHisProGlnPr 977
781 GCTCGGTTCTAGAGGCTCTTGCAACAA..ACGGCC..... 745
977 oThrGlnValGlnPheValCysAspArgGlnGluThrAlaAlaSerLeu 994
744 .....AATGTAATACATCTGATTAATGAT 718
994 alThrSerAlaValGlyGlnGlnAsnGlyAsnValAlaGlyValCysSer 1010
717 GGTCCACACGGTTTATTCGCGCGCAGCGCTCGATAAATGATGTGC 668
1011 AsnProPro.....CysGlnThrHisGlyThrThrThrAsnThrAl 1025
667 TGCCATCTCAACCGGAGATGCGCGCGCGGATCATGTGT..... 625
1025 aThrThrAlaThrSerAsnMet..AlaGlyGlnHisGlyCysSerAsn 1041
624 .....TTGCATGTTGGACATTTTGCACGCGACGTTCTGC 589
1041 roProCysGlnThrHisGlnThrGlyThrThr..SerThrAlaThrThrAl 1057
588 GCGAGCTGCTTACAAACATGATTTGGTGGGTGAACGCGCTCAATA 539
1057 aMetSerSerMetGlyThrGlyGlnGlnAlaArgAspThrAlaArg...Th 1073
538 CCAACAGGCGG.....GTTTGAATCTCGCGCGCTTCTTGATA 498
1073 hrAsnThrProThrValAlaArgIleThrValAlaProGlyAlaLeuGlu 1089
497 ATGACCGTAGGTCGCGCA.....GCCAGCGGATTT 469

```

```

1090 ArgValGlnGlyThrValValProGlnCysGlnThrGlnGlnThrAsnMe 1106
      |||||:
468 GGTGCCATTCGATTCGACGAAATGCGGAACGGCTCGCATCGCGCAG 419
      |||||:
1106 tThrThrThrMetThrValGlnAlaThrGlnAlaProCysSerAlaG 1123
418 GAATTTGCTGACGAGGAGGCTG.....CGCAGCGGATCCCAACACG 375
1123 lProLeuLeuAlaGProSerValAlaLeuGlnSerIleHisSerPro 1139
374 GATTGGATCAGGTTGCGCGCGCTTCTCGCCCTTACGTTAGTTGCCGCGC 325
1140 AlaPheValGlnLeuAlaLeuProSerValAlaGlyLeuSerGlyPr 1156
324 TTCAGTGGCGTACCGTTCAACATCGATTTGCTGCTTCCTTCACGCGCA 275
1156 oSerSer.....LysAspMetProThrGlyArgGlnProGlyThrTy 1171
274 TCACGACTGACTGAGTACGCGCGCTTTCGCAAGCGGTGATTCGCGCGCAT 225
1171 lThrTyThrThrAsnThrProThrThrThrArgSerIle...MetVal 1186
224 TTGCTGAGACCGCGCGCAGTAAACACACCGCGGATTTCTTTGTCTTC 175
1187 AlaGlyGlyLeuGlyAlaAlaAlaArgValAlaPro..... 1197
174 AACACACACTTGAGCTTTTTCAGCGCA..TCGCTTCCTTGACTTCA 128
1198 .ThrSerThrTyThrGlnSerLeuGlnAlaSerProSerSerThrMet 1214
127 TCGAGGGGCGCATACCGGATATCTTCGCGCAAC..... 93
1214 hMetThrAlaLeuGlnAlaLeuLeuCysProSerAlaThrValThrGln 1230
92 .....AA 91
1231 ValCysSerAsnProProCysGlnThrHisGlnThrGlyThrThrAsn 1247
90 CGGAGCTTGGTAAATGCCGCGCGCTCGTAAACGCGTGTCCGCTGCG 41
1247 rAlaThrThrSerAsnAlaGlySerAlaGlnArgValCysSerAsnPro 1264
40 CC 39
1264 to 1264
seq_name: sp_mammal:029071
seq_documentation_block:
ID 029071 PRELIMINARY; PRT; 528 AA.
AC 029071:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE GASTRIC MUCIN (FRAGMENT).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9825;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE=GASTRIC MUCOSA;
RX MEDLINE=95275264; PubMed=775593;
RA Turner B.S., Bhaskar K.R., Hadjopoulos-cladaras M., Specian R.D.,
RA Lamont J.T.,
RT "Isolation and characterization of cDNA clones encoding pig gastric
RT mucin."
RL Biochem. J. 308:89-96(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE=GASTRIC MUCOSA;
RX MEDLINE=94102478; PubMed=7506218;

```

Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
 LaMont J.T.:
 "Pig gastric mucin: isolation and characterization of a cDNA clone
 with a novel tandem repeat."
 Gastroenterology 106:200-200(1994).
 RL EMBL; U10281; AAC48526.1; ..
 DR NON_TER 1 1
 FT 528 528
 SEQUENCE 528 AA; 49907 MW; 0BFOF6879203B2EA CRC64;

alignment_scores:

Quality: 113.50 Length: 458
 Ratio: 0.493 Gaps: 19
 Percent Similarity: 50.218 Percent Identity: 20.961

alignment_block:

US-09-303-518d-125/rev x Q29071 ..

Align seg 1/1 to: Q29071 from: 1 to: 528

```

1343 CAGCTTCCTTCCTCAATG.....GTTCCAGCACTTTCGCAACAG 1303
      |||||  |||
149 GlnProSerSerSerSerProIleSerSerThrVal..... 162
1302 CGGCGCTATTCGTAATTCGCGCGGACGACGAGCTGCACAAAGGAGGT 1253
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
163 .....SerValGlnProSerSerSer.....SerAlaProT 174
1252 CTCTTCCTTCCTTCCTCAATG.....GTTCCAGCACTTTCGCAACAG 1206
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
174 ThrThrSerAlaThrSerValGlnProSerSerSerSerProIle 190.
1205 TCGCGCGATTAATTCGCGCAACGAGGTGGGAGATATTCACAGG 1156
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
191 SerSerThrValSerValGlnThrSerSerSerSerSerValProThr 207
1155 CATACCGCGCTCAATGACCAATGCGGACCATGCGCGCGCT 1106
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
207 rSerThrThrSerValGlnPro.....SerSerSerSerSerValProT 222
1105 TGACGGCGTGTGACTGAGAGATTTGTTTTCAGAAATGCGCGAG 1056
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
222 ThrThrSerAlaThrSerValArgSerSerSerSerSerThrPro... 237
1055 GTTTCAGCGGTGATGAGATTTTTCGCGCTGCGCGCAACCCAGCGGA 1006
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
238 .....IleProSerThrThrSerValGlnProT 247
1005 CAGCTTCCTTCCTTCGCGCT...TCTTCGATAAGGAAATCTGATGCT 959
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
247 rSerSerSerSerAlaProThrThrSerAlaThrSerVal..... 260
958 AGCGTCCCAATTAATCGTGCCTGCTGTGTAATCGCGCGCTCATAC 909
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
261 ..GlnProSerSerSerSerSerThrProIleProSerThrThrSerVal 276
908 GACCGGAAATCAGCGGTTCTGTGTGACCAATTCGCGCGAGTAAT 859
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
277 GlnPro.....SerSerSerSerSerAlaProThrThrSer 288
858 TTGCGATACCTTCGACCCCAACGATACGCAAGAGCGGTTGTTGA 809
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
288 rAlaThrSerValGlnProSerSer.....S 297
808 CTTGAGAACCACTAGGCAATCAGCGCTGCGGTTCAGACGCGCTGT 759
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
297 rSerSerProIleSerSerThrIleSerVal.....GlnProSer 311
758 GCAACCAACGCGCAATGTAATTCATCTGATTAATGATGATGCACAC 709
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
312 SerSerSerSerSerProThrThrSerThrThrSerValGlnProSer 328
  
```

```

708 GGTTTTATTCGCGCGACCGCTCATGAATGAATGTCGCTCCACTCA 659
      |||||  |||
328 rSerValSerAlaProThrThrSer.....AlaThrSerValG 341
658 AACCGGAGATTCGCGCGCGCAATTCATGTTGTTGATGGACGA 609
      |||||  |||  |||
341 InProSerSerSerSerProIleSerSerThrIleSerValGln 357
608 TTTTCAGACGCGCTTCGCGCGCTGCTTACCAATGATTTTCG 559
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
358 ProSerSerSerSerSerProThrThrSerThrThrSerVal.... 372
558 TTCGTCAAACGCGCTCAATACCAACAGCGCGCTGTAATCTCGCG 510
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
373 .....GlnProSerSerSerSerValGlnProSerSerSerVal 381
509 .....GCTTCTTGAATGACCGCTAGCGGTGCGCGCGATTTG 465
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
381 rThrThrSerAlaThrSerValGlnProSerSerSerSerValPro 397
464 TCCATGCAATTCGACGAGATGCGCAACGCTGCGCATCGACGAGAT 415
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
398 ThrThrSerAlaThrSerValArgSerSerSerSerSerThrPro 414
414 TTTGCTGAACGAGCGGTGCGCGCGCATCACAACCGGATTCGATCA 365
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
414 eProThrThrThrSerValGlnProSer..... 423
364 GGTTCGCGCGCGCTTCGCGCGCTTAACTTTCGACGCGCTTCAGT 315
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
424 .....SerSerSerSerValProThrThrSerAlaThrSerVal 436
314 TAGCTTCGAATTCGATTCGTCG...TTCCTTCAGCGCAATACAGC 268
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
437 GlnThrSerSerSerSerSerThrProIleProSerThrThrSerVal 453
267 TGACTGAAGTACCGCTTCGCGCAACGATGATTCGCGGATTTGCTG 218
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
453 nProSerSerSerSerSerAlaPro..... 461
217 AAGCGGCGCATTAACACACCGCGCGATTCCTTTGCTTCGAACAGC 168
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
462 .....ThrThr.....SerAlaThrSer 467
167 ACTTGCCCTTTTTCAGCGCATGCGCTTCCTTCGATTCATGAGAGG 118
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
468 ValGlnProSerSerSerSerSerProIleSerSerThrIleSerVal 484
117 CATACGCGCATTCCTTCGCGCAACGAGCTTCGCTGATG..... 75
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
484 GlnProSerSerSerSerSerProThrThrSerThrThrSerValG 501
74 .....GCCGCGCGCTGTAACGCGCTTCGCGCTG 42
501 InProSerSerSerSerSerAlaProThrThrSerAlaThrSerValG 517
41 CCGCGCATGCGCAGCTTACACCT 18
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
518 ProSerSerSerSerSerProPro 525
  
```

seq_name: sp_bacteria:0935J8

```

seq_identification_block:
ID 0935J8 PRELIMINARY: PRT: 575 AA.
AC 0935J8,
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE USC3-5P.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cytophasterineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
  
```

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CECT 422;
 RA Poza M., Sileiro C., Villa T.G.;
 RT "Clone USC3 from *Mycococcus xanthus* CECT 422 strain."
 RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases
 DR EMBL; AY033403; AKK49004.1; -
 SQ SEQUENCE 575 AA: 62586 MW; D2L19B7G363C55634 CRC64;

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 112.00 | Length: | 573 |
| Ratio: | 0.521 | Gaps: | 29 |
| Percent Similarity: | 37.522 | Percent Identity: | 21.815 |

```
alignment_block:
```

US-09-303-518D-125 x Q93SJ8

Align seg 1/1 to: Q93SU8 from: 1 to: 575

```

9  AATCAAAAAGCTTAACCTGCCCCATCCGGCGGCGACGACCGAGCAAC.. 56
10  :ATC :||| :||||| :||| :||| :||| :||| :||| :||| :|||
11  31  ***Ala***ArgMet***ProAlaSnAlaHisIuPheSerAlaAla** 47
12  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13  57  .....CGTTTACGACGGCGCCGCCCATATTACCGAAGTCGCGTTG 93
14  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
15  47  *GlyAlaArg**ArgProGlnAlaProAla***ArgSerAlaAlaIle 64
16  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
17  94  CTTGG.....CGAAGAATATGCGCGGTATGCGCCCTC...GATGAA 131
18  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
19  64  IaSerThr***ProArgSer***AlaLysAlaGlyThrLeuAlaThrPro 80
20  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
21  132  AGTCAGGAAGGCGGATCCGCTCAAAAAGGCGCAAGTCTGTTGAAGACA 161
22  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
23  81  ThrHisGlySerSerAlaArgSerLysArgProAsnPro.....ArgI 95
24  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
25  182  AAAAAGATCCGGCGCGTGTCTTACTGCGCGGCTTCAGCAAAATCGCC 231
26  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
27  95  ngInArGArgGlyAlaArgSnAla.....SerArgA 105
28  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
29  232  GCGATTTCACCGTGGCGAAAGCGCGTACCTCACTAGTGTGATTCGGT 281
30  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
31  105  Iaasp.....ArgGlyAlaLeuArgAlaGlyLys... 114
32  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
33  282  TGAAGCGACGACGAAATCGAGTTTGAACGCTGACGACCTGGAAGCGCTG 331
34  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
35  115  .....AlaAlaGlyThrAlaArgValProProIaI 125
36  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
37  332  CAAACTTAAGCGCGA..... 347
38  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
39  125  yProIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 142
40  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
41  348  .....AGAAATGCGCGCGCAACCTGA.....TCCAAATCCGGT 379
42  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
43  142  rGcGsglyThrArProIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 158
44  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
45  380  TCGGACGCTCGCTGCGCACCCGCTCCGTTACGAAAA...TTTCGCCGTC 426
46  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
47  159  AlaaGlySerAlaProIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 475
48  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
49  427  GATGCGG.....AGCGTTGCGGCATCTTCGTAATGC 458
50  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
51  175  galIaProIeugInaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 191
52  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
53  459  GATGACACCAATCCGCTGCGCGCACCCACCAAGGT..... 494
54  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
55  192  ProAlaAlaArgCysProGlySerCysProProGlyAlaIaIaArgYarSe 208
56  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
57  495  .....CATTTCAAGAACGCGCGCGAGATTTCAAAAGCGGCGCTG 534
58  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
59  208  rTTPrometGInTyIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 224
60  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
61  535  TTGGAATTAGACCGTTTGACCGAAGCGCAAAATTCATGATTTTAAAGGAC 584

```

```

225 .....ProCys..... 228
226 ..|||..|||||..
227 ..ArgLeuProAlaProProGlyProCysGlnAldArg..... 238
228 .....
229 .....AlaAlaSerAlaValAlaArgTrpHisArgHisAlaGlnCysP... 674
230 .....
231 .....TCATTTCAT... 683
232 rolaIalgaIgrTrpaIGserAlaTrgHlraIProlaIdgaISerValnHisThr 268
233 .....
234 .....CGAGCCGGTGGCGCGCAATA 703
235 ArgAspTrpHiseraIalaCysSerProMetArgAlaTrpHlraIrgAIdgas 285
236 .....
237 AAcCGGTGGAGCCATCAATTATCA..... 728
238 nserasnclurHlraIProlleSertTrpnaCysglYThSeCysglur 3020
239 ..AGATGAATTTACCATTGCCCGTTGTCTTGCACAAGCCGCTGAAAC 776
240 roaIaCysAlaIaIaSetrPrOCysAlaCystRylpsASnsenArpadg 318
241 CGAGCGCGTGAATGCCCTCAGGTGGTGTCCAAGTCACAAACC GGCTCT 826
242 LysASNdrAgLeuAlaSerAlaValaAlaSNprOsetrThrasnclurAlaIdre 335
243 TGCGGAC..... 833
244 uSerMetProLysProGlyIleAlaASNdrValaGlnArgGlnAlnArgS 352
245 eArgInISalateThLeuASNproIdrgLysGlnArgGlnAlaIdCysArg 368
246 AspArgLyuArnArgGlnGlnIygluArgArGlnAlaValLeuAlaAs 385
247 .....GAAAGTATCGCAATTA CTGCGGGGCAATGGTTGACACAGAC 888
248 rArgASrGlnAlaValaIaIdIuarGrHrGlnArGasrPaAlaIgLnL 402
249 AACCGCGTGAATTCGGGTTCCGATTGAACGGCGCGCATTCACAAGCGGC 938
250 euPrOGluArgValaIghlaIdrgGlyMeIeudArGmerProSerLeuArgSn 418
251 GCACGATTATTTGGAGCGTACACACAAATAGACTTCGGTTCCGAAGAAG 988
252 Ala.....ProGlyAlaGlnHisGlnCysLysAlaAr 429
253 GCGGCAACAAAGCGTGTTCGGCTGGGCTGGTTCGCGCGCACCGACAAATAC 1038
254 gPrOAlaIalagLyLeuIlleArgIleuAsrPalagLyProAlaGlyAlaLeuN 446
255 TCACTACAGCGTACACCGTCGCGCATTTCCGTGAAMAAACAACCTTTCAA 1088
256 IsgInPheTrpHlraGln***ArgPro..... 454
257 GTTCAACACACAGCGGTACACGCGCGGACCGCGCGCATGGTGCAGATTGTA 1138
258 ...GlnHisGlnAlaGlnArGrHrGlnArGlnArGrProAlaProAr...Hl 469
259 CTTACAGCGCGTGAATGCCCTTGATATCTGTGCCAACCCCTCTTTGGCGC 1188
260 salarIdgrHnArg.....AlaA 475
261 GATTATATGTGTGGCGATACCGACAGCGCGAGGCATTGGG...TTGGCTT 1235

```



```

297 lYthSerThrThr.....ThraSerSerCluser:307
158 TTTTGGAGCATGCGCTTCTTCTTCAATCAGGCGGCGCATACCGGC 109
108 ATATCTTCG...CCAGCAACGCGACTGCGTAATGCGCGG.....69
314 lYtSerSerMetProPhSerThrThrThlValThlSerGlySerAlaI 331
68 .....CGGTGTAACGCGCTTCTCGCGTCCGC 39
331 lelleProAspHisAsnGlySerSerValLeuPro 342

seq_name: sp_invertebrate:Q9U0V2

seq_documentation_block:
ID Q9U0V2 PRELIMINARY; PRT: 2207 AA.
AC Q9U0V2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE POSSIBLE MUS308 HOMOLOG (FRAGMENT).
GN L8342.01.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_Taxid=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Masny D., Purrelle B., Goffeau A., Ivens A.C., Lawson D., Quail M.,
RA Rajendream M.A., Bartell B.G.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.;
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL EMBL; AL122012; CAB58415.1;
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00270; DEAD.1.
DR Pfam: PF00271; Helicase_C.1.
DR SMART: SM00490; Helicase.1.
KW ATP-binding; Helicase.
FT NON_TER 2207 2207
SO SEQUENCE 2207 AA; 229262 MW; E461565DA0017B0C CRC64;

alignment_scores:
Quality: 112.00 Length: 457
Ratio: 0.599 Gaps: 24
Percent Similarity: 40.919 Percent Identity: 22.976

alignment_block:
US-09-303-518D-125/rev x Q9U0V2

Align seg 1/1 to: Q9U0V2 from: 1 to: 2207

1232 CAACCAATGCGTGGCGCTGTCGTAATCCGCGATTAATCGGCGCA 1183
1183 GlnProArgLeuGlySerGlnLeuSerValAlaThrThr.....824
1182 AAGAGGGTGGCGAGATATCCAGGCGATCAGCGC.....T 1145
825 .....ThleGlnAlaSerAlaGlySerAlaArgAsnValAlaAla 839
1144 CGTAAGTACCAATCGCACCATGCGCGGCGCGCGCTTGAGCGGTG 1095
1095 erValAlaGlnProAlaThrProValSerSerProSerGlnAla... 854

1094 TTGAAGTGAAGAGTTTGTTTTTCAGCAATGGCCGAGGTTGTACGCT 1045
854 .....
1044 GATGAGATATTTGTCGGGTGCGCGACCGACCGACAGTCTTTGC 995
855 .....CysAlaThrThrHisPro...ProAspLeu 864
994 TCGCGCTTCTTGCATACGAAATCTGATGTGTAGCGTCCCAATAA 945
864 rPluProSerLeuProThrGlnHis..... 872
944 TCGTGGCGCTGTGTATCGCGCGTTCAATACCGAATACATAC 895
873 ValSerProProCys.....AspProLeuLeu 882
894 GCGGTGTCTGTGTCACCAATTCGCCGCGAGTAATTCGATCTTGC 845
882 lArgValSer...SerGlnGlyGlyAlaGlnGlyCys..... 894
844 CACCAAAACGTACGAGAGCGCGGTTTGTGACTTGAGACCACT 795
894 .....
794 AGGCAATACAGCGCTGCTGTTCAGACGCGCTTGCAACACAGCGC 745
895 .....AlaAlaGlnSerValPheArgMetGlyValAlaAlaProThrPr 909
744 ATGTGTAATACATCTTGATTAATGATGTGTCACACGGTTTATTCGCGC 695
909 cThrSerLeuGlySer.....AspValLeuSerSer 920
694 CGACGCGCTCGATGAATGAATGCTGTCACATCAACCGCAGGATGC 645
920 lAthrGly.....ValSerAlaLeuSerAlaAlaAsnAla 931
644 GGGCGCGCAATTCATGCTTTTCATGTTCGCA..... 612
932 AlaProProLeuSerAspLeuHisValAlaThrAlaLeuProTyrLeuAlaThr 948
611 .GCATTTTCAGACGCGCTGCGCGCGCT.....GCCT 578
948 rAlaAlaAlaGlnGlySerGlyAlaProAlaValProAlaArgProAla 965
577 TCAACACATGATTTGCTGTCGTCGTCACACGCTCAATACACAGCGCG 528
965 rGThrCysPheThrLeuHisSerAlaAlaArgThrThrGlyThr..... 979
527 GCTTGAATCTCGCGCGCTTCTTG.....ATAATGAC 493
980 .....LeuSerSerSerAlaAlaAlaThrSerThrGlnGlnProThrThrAl 995
492 CGTAGGTCGCGACGCGACGATGTCATTCATCGCATTCAGCAGAGTGG 443
995 aAlaAlaSerAlaSerThrGlyAlaValProProAlaLeuThrGlnAla 1012
442 CG..... 441
1012 erTyrLeuSerAlaGlnGlyCysPheArgGlyGlyThrAlaGlnLeu 1028
440 .....AAGGCTCGGCGATCGACGCGCA..... 417
1029 GlnGlnLeuProGlnHisGlyAlaAlaAlaSerProGlnGlyThrGlnGlyLe 1045
416 .....ATTTCGTGACGAGCGGTGCGCGACGCGACGTC..... 384
1045 uMetValSerLeuAspGly...ValAlaSerAlaThrProSerAspThr 1061
383 .CACAAACGGAATGATGATGCTGCGCGACATCTTCGCGCGCTTAC 336
1061 rGlySerMetProSerSerMetGlnSerArgHisThrThrGlyCysAlaGln 1077

```


01-MAY-2000 (TREMUREL. 13, Created)
 01-MAY-2000 (TREMUREL. 13, Last sequence update)
 01-DEC-2001 (TREMUREL. 19, Last annotation update)
 RNA BINDING PROTEIN (FRAGMENT)
 Homo sapiens (Human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 Ohtaki S., Umeki K., Sawada Y.;
 "Homo sapiens mRNA for RNA binding protein, partial cds."
 Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 EMBL: AB016087; BAA83713.1;
 DR InterPro: IPR002965; P_Rich_extensn.
 DR PRINTS: PRO1217; PRICHTEXTNSN.
 FT NON_TER 1 1262
 FT NON_TER 1 1262
 SQ SEQUENCE 1262 AA; 138003 MW; 75E76BD73FA14295 CRC64;

alignment_scores:
 Quality: 111.50 Length: 435
 Ratio: 0.578 Gaps: 20
 Percent Similarity: 44.368 Percent Identity: 23.678
 Alignment_block:
 US-09-303-518d-125 x 09040 ..
 Align seg 1/1 to: 09040 from: 1 to: 1262

18 AGGCTTAACCTGCGCATCGGGGAGAGCAAGCGTTTACGACG 67
 |||||
 273 ArgSerArgSerProAlaThrAlaLysArgLysArgSerArg 289
 68 GCCC.....GCCATTACCGAAGTCG...GTTGCTGGCGAAGA 105
 |||||
 289 rProthrLysArgLysHisSerArgSerArgSerProGlnTrpArg 305
 106 TATCCCGGTATGCGCC.....CTCGATGAAGTCAAGAGGCGG 146
 |||||
 306 ..SerArgSerAlaGlnArgTrpLysArgSerArgSerProGlnArg 321
 147 TGCCGTCAAAAAGGCCAAGT..... 167
 |||||
 322 GlyArgSerArgSerProGlnArgProGlyTrpSerArgSerArg 338
 168GCTGTTTGAAGCAAAAAGATCGGCG 195
 |||||
 338 rGlnArgArgLysArgSerArgSerAlaArgLysArgSerHis 355
 196 GTGGGTATTACGCGCGGCTTACGCAAAATCGCGGATTCACG 245
 |||||
 355 rSerProAlaThrArgLys.....ArgSerArgSerArgTrp 369
 246 CGAAAAGCGCGTACT.....TCAGTCAG. 268
 |||||
 370 ArgArgGlyArgSerArgSerArgTrpProAlaArgArgSerArg 386
 269 TCGTGATTCGCGTGAAGCAAGCAAGATCGATTGAAGCGTAC 318
 |||||
 386 rArgTrpProThrArgArgArgSerArgSerArgTrpProAla 403
 319 CCGAAGCGCGTGAAGCAAGTGAAGCAAGTGCAGCCCAACTG 368
 |||||
 403 LysArgSerArgSerArgTrpProAlaArgArgSerArgTrp 419
 369 CCAATCGGTTTGGACTCGCGTCCGACCCGTC...CGTTACAG 414
 |||||
 420 ProValArgArgSerArgSerArgSerArgTrpProAlaArg 436
 415 ATTCTGCGGTGATGCGGCGGCTTCCGCACTTCGTCATTCG 464
 |||||

436 gSerArgSerArgTrpProAlaArgArgLysArgSerArgSerArg...T 452
 465 CACCAATCCGCTGCTGCGCGACCGCTACGCTATATCAAAAGCCGCG 514
 |||||
 452 hrProAlaArg.....ArgGlyArgSerArgSerArgTrpProAla 465
 515 AGGATTCAACGCGCGCTGTTGATTTAGCCGCTTGGACGCAAGCA 564
 |||||
 466 ArgArgSerArgLysArg.....SerArgSerArgTrp 475
 565 ATCCATGTTTGAAGGACGCTGCGCGACGCTGCGCTGTAATGCTGC 614
 |||||
 475 hrProAlaArgArgLysArgSerArgTrpProAlaArgArgLys 491
 615 CACCATCAAGCAACATGATTCGCGCGCGCGATCTCGCTTGAAGTG 664
 |||||
 492 SerArgSerArgSerLeuValArgArgLysArgSer..... 503
 665 GCACGCGCATTCATTTATCGAGCGGT.....C 693
 |||||
 504 HisSerArgTrpProGlnArgArgLysArgSerArgSerArgLys 520
 694 GCGCGCAATTAACGCTGCGCGACCATCATTTCAAGATGTAATAC 743
 |||||
 520 rGlySerArgSerArgTrpSerGlnArgArgSerArgSerArgSer 536
 744 TGCCGCTGTTTGAAGCAAGCGCGCTCGAAGCAAGCGCGGTAT... 788
 |||||
 537 ProGlnMetLysSerArgLysSerArgLysSerArgSerArgSer 553
 788 788
 553 rSerProArgSerLysAlaLysSerArgLysSerLeuArgArgSer 570
 789TGCCATGAGTGGTTCACAGTCACAAACCGCGCT 824
 |||||
 570 erGlySerSerProGlyProLys.....GlnLysSerGlnTrpPro 584
 825 CTTCGCTACCGTGGTGGCGAAGATTCGCAATTTACGGGGCGAT 874
 584 584
 875 TGGTTGACACAGCAACCGCGGTATTCGCTGCTGATTTGAACGCG 924
 584 584
 925 ATTACAGAGCGCGACGATTTATTTGGACGCTACCAATCAGATTTC 974
 |||||
 585ArgArgSerArgSerGlySerSerGlnProLysAlaLysSer 598
 975 CGTATTCGAAGAGCGCGCAAGCAAGGCTGCGGCTGGGTTGCGCGC 1024
 |||||
 599 ArgTrpProProArg.ArgSer.ArgSerSerSerProProProLys 614
 1025 AGCCGCAAAATACCTCATCAGCGGTACAAACCTCGGCCATTTCC 1074
 |||||
 615 GlnLysSerLysTrpProSerArg.....G 623
 1075 AACCAACTCTCAAGTTCAACACAGCCGTCACAGCGCGCGCACCG 1123
 |||||
 623 nSerHisSerSerSerProHisProLysValLysSerGlyTrpPro 639
 seq_name: sp_bacteria:p94909
 seq_documentation_block:
 ID P94909 PRELIMINARY; PRT; 529 AA.
 AC P94909;
 DT 01-MAY-1997 (TREMUREL. 03, Created)
 DT 01-MAY-1997 (TREMUREL. 03, Last sequence update)
 DT 01-NOV-1998 (TREMUREL. 08, Last annotation update)
 DE HYPOTHEICAL 38.7 KDA PROTEIN (FRAGMENT).
 OS Microbacterium ammoniaphilum.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OK Actinomycetales; Micrococccineae; Microbacteriaceae; Microbacterium.
 NCBI_TaxID=42460;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15354;
 RX MEDLINE=96257250; PubMed=8654988;
 RA Striebel H.M., Seebler S., Jarsch M., Kessler C.;
 RT "Cloning and characterization of the MamI restriction-modification
 RT system from Microbacterium ammoniaphilum in Escherichia coli.";
 RL Gene 172:41-46(1996)
 DR EMBL: X79027; CAA55649.1; -
 KW Hypothetical protein.
 FT
 NO_TERM 1
 SO SEQUENCE 529 AA; 58716 MW; 5DC1216237DF772D CRC64;

alignment_scores:
 Quality: 111.00 Length: 526
 Ratio: 0.491 Gaps: 29
 Percent Similarity: 42.966 Percent Identity: 24.715

alignment_block:

US-09-303-518D-125 x P94909

Align seg 1/1 to: P94909 from: 1 to: 529

```

21 TCCTAACCTGCCATCGCGGCGAGACCGGCAAGCGCTTACGA...CG 67
   ::::::::::::::::::::
49 AAlaySProlayarglaiglylproarglylthgltlelproarg 65
   ::::::::::::::::::::
68 GCCCGGCAATACCGGCAAGTCCGCTGCGCGCAAGATATGCGGTAG 117
   ::::::::::::::::::::
65 gglYserhNstYarghNstYarghNstYarghNstYarghNstYarg 77
   ::::::::::::::::::::
118 CGCCGCTCGATGAAGTCAAGGAGCGGATGCGGCAAGGAGCGGCAAGT 167
   ::::::::::::::::::::
78 .....ArgGlnAlaHisProleu 83
   ::::::::::::::::::::
168 GCTGTTGAAGACAAAGATCCGGGCGTGTCTTACTGCGCGCGCTT 217
   ::::::::::::::::::::
84 GlnATPronHsGlnGlylproarglylthgltlelproarg 98
   ::::::::::::::::::::
218 CAGGCAAAATCGCGGCAATGCGGCGGCAAGGAGCGGCAAGGAGCGGCA 267
   ::::::::::::::::::::
98 lArgSParsarglaiglylthgltlelproarglaiglylthgltlelpro 115
   ::::::::::::::::::::
268 GTCGATGTCGCGT..... 281
   ::::::::::::::::::::
115 lArgSParsarglaiglylthgltlelproarglaiglylthgltlelpro 131
   ::::::::::::::::::::
282 .....TGAAGCAAGCAAGGAGCGGCAAGGAGCGGCAAGGAGCGGCA 322
   ::::::::::::::::::::
132 GlnAlaHisProleuArglaiglylthgltlelproarg 141
   ::::::::::::::::::::
323 AAGCGCTCGCAAACTTAAGCGG.....CGAAGAGTGGCGGCGGCAAGT 366
   ::::::::::::::::::::
142 AAlaySProlayarglaiglylthgltlelproarglaiglylthgltlelpro 158
   ::::::::::::::::::::
367 ATCCGATCGGCTTGTGATGCGGCGGCAAGGAGCGGCAAGGAGCGGCA 416
   ::::::::::::::::::::
158 lSProlayarglaiglylthgltlelproarglaiglylthgltlelpro 173
   ::::::::::::::::::::
417 TCGTGGCTCGATGCGGCGGCAAGGAGCGGCAAGGAGCGGCAAGGAGCGG 442
   ::::::::::::::::::::
174 .....ArgATGarglaiglylthgltlelproarglaiglylthgltlelpro 188
   ::::::::::::::::::::
443 CCAGCTTCGCAATGCGGCAAGGAGCGGCGGCAAGGAGCGGCGGCAAGGAG 488
   ::::::::::::::::::::
188 gHis.....GlySParsarglaiglylthgltlelproarglaiglylthgltlelpro 203
   ::::::::::::::::::::
489 .....TAGGTCATTAAT.....CAAGA 506
   ::::::::::::::::::::
203 SProlayarglaiglylthgltlelproarglaiglylthgltlelproarg 219
   ::::::::::::::::::::

```

```

507 AGCCGC.....CGAGATTTCAACGGCGCTGTGCTATTA 544
   ::::::::::::::::::::
220 AAlaySProlayarglaiglylthgltlelproarglaiglylthgltlelpro 234
   ::::::::::::::::::::
545 GCCGTTTGAAGCAAGGAGCGGCAAGGAGCGGCAAGGAGCGGCAAGGAGCGG 590
   ::::::::::::::::::::
235 .ProAlayarg.....HisProAlayarglaiglylthgltlelproarg 249
   ::::::::::::::::::::
591 .....AGAGTGGCGCTGCAAGTGC 611
   ::::::::::::::::::::
249 rGlaArgGlnGlyAlaArglaiglylthgltlelproarglaiglylthgltlelpro 265
   ::::::::::::::::::::
612 TGCCAGATCGCAACACATGATCGGCGGCGGCAAGGAGCGGCAAGGAGCGG 661
   ::::::::::::::::::::
266 LeuGlnArgarg.....GlyArgAlaSProlayarglaiglylthgltlelpro 279
   ::::::::::::::::::::
662 GTGCGAGCGCAATGATTCATCGAGCGGCT.....CGGC 696
   ::::::::::::::::::::
279 nGlyArgAlaArgarglaiglylthgltlelproarglaiglylthgltlelpro 296
   ::::::::::::::::::::
697 GCGATTAAGGCGGCTGAGCAATGATTCATCGAGCGGCAAGGAGCGGCAAGGAG 746
   ::::::::::::::::::::
296 rGln.....AspArgAlaArglaiglylthgltlelproarg 304
   ::::::::::::::::::::
747 CCG.....TT 751
   ::::::::::::::::::::
305 ProArgProArgGlyAlaArgGlnAlaArglaiglylthgltlelproarglaiglyl 321
   ::::::::::::::::::::
752 TGCTTGCAGAGCGGCTGCAAGGAGCGGCAAGGAGCGGCAAGGAGCGGCAAGGAG 797
   ::::::::::::::::::::
321 yAlaGlyAlaArgProAlaArgProArgGlyAlaArgGlnAlaArglaiglylthgltlelpro 338
   ::::::::::::::::::::
798 .....TGCTTGCAGAGCGGCTGCAAGGAGCGGCAAGGAGCGGCAAGGAGCGG 836
   ::::::::::::::::::::
338 lAspLeuProHsGlnGlylproarglaiglylthgltlelproarglaiglylthgltlelpro 354
   ::::::::::::::::::::
837 TTTGGTGGCAAGGATTCGCAAGGAGCGGCAAGGAGCGGCAAGGAGCGGCAAGGAG 886
   ::::::::::::::::::::
355 GlnGlyLeuGln.....GlyAlaArglaiglylthgltlelproarg 365
   ::::::::::::::::::::
887 ACAAGCGGCGGATTCGCGGCGGCAAGGAGCGGCAAGGAGCGGCAAGGAGCGGCA 930
   ::::::::::::::::::::
365 gHisProHsArglaiglylthgltlelproarglaiglylthgltlelproarg 382
   ::::::::::::::::::::
931 CAGAGCGGCGGATTCGCGGCGGCAAGGAGCGGCAAGGAGCGGCAAGGAGCGGCA 980
   ::::::::::::::::::::
382 euArgarg.....ProArgArgargarglaiglylthgltlelproarg 391
   ::::::::::::::::::::
981 CGAAGAGCGGCGGCAAGGAGCGGCGGCAAGGAGCGGCAAGGAGCGGCAAGGAGCGG 1030
   ::::::::::::::::::::
392 ArgArgGlyProArgGlnGlyAla.....GlyAlaHisGly 404
   ::::::::::::::::::::
1031 ACAATATCTCATGAGCGGCGGCAAGGAGCGGCGGCAAGGAGCGGCAAGGAGCGGCA 1080
   ::::::::::::::::::::
404 yGlnGlyLeuArgHsAlaArgargarg.....ArgGlyAlaProSerGln 420
   ::::::::::::::::::::
1081 CTCTTCAGATTCAGAGCGGCGGCAAGGAGCGGCGGCAAGGAGCGGCAAGGAGCGGCA 1130
   ::::::::::::::::::::
421 LeuAlaValLeuArgGlnGlyAlaGlnAlaSerAlaHisLeuAla 437
   ::::::::::::::::::::
1131 GATGTCATTCAGAGCGGCGGCAAGGAGCGGCGGCAAGGAGCGGCAAGGAGCGGCA 1180
   ::::::::::::::::::::
437 uTPMetHsSerGlyArgleu.....LeuArgGlnLeuAlaGlyLeu 452
   ::::::::::::::::::::
1181 TTTTGCAGATTCAGAGCGGCGGCAAGGAGCGGCGGCAAGGAGCGGCAAGGAGCGGCA 1212
   ::::::::::::::::::::
452 euCysArgSParsArgValGlnSerHsLeuArgValThrThrArgAla 468
   ::::::::::::::::::::
1213 .....AGCGGCAAGGAGCGGCGGCAAGGAGCGGCGGCAAGGAGCGGCAAGGAGCGGCA 1250
   ::::::::::::::::::::
469 AspAlaSerLeuSer ArgArgThrThrValMetAsn...ThrArg 484
   ::::::::::::::::::::

```


521AATCTCGGCGGCTTCTTGAAATGACCGTAGCGCA 480
 |||||
 1471 euAlaProlysrhrlleYrProSerSerleuGlnThrGlySerSer 1487
 479GCCAGCGGATGGTGTCATCGCATGACGAGCAT 445
 |||||
 1488 ProThrProProGlyThrSerSerleuValAlaGlySerArgAl 1504
 444 GCGGACGGGCGCGCATCGACGAGAAATTTGCTGAACGACG... 399
 |||||
 1504 aSerSerAsnTYrProThrThrAlaSerlleuLuhhrspgLySerSerg 1521
 399
 1521 luGlInglIngluGlInuAsnThrArglleLeuProgluGlInhrLeuSer 1537
 398 ...GTGCGCAGCGCATCCCAAAACCGGATTTGATCAGTTGCGCGCAG 352
 |||||
 1538 ProAlaSerAlaArgLeuTrpThrProAspMetlleSerLySTyrlleSe 1554
 351 TTCTTCGCGCTTAAGTTTGCAGCGCTTCAGTGCAGTGCCTCAACT 302
 |||||
 1554 rSerLyAlaIleSerThrGlnSerValSerSerAlaGlnProserT 1571
 301 CGATTTTCGTG...TTGCTTCAACGCGCATCAGACACTGAGTAG 255
 |||||
 1571 yrlLeuSerSerAsnLeuProGlnSerSer..... 1581
 254 CGCTTTTCGCGCATCGGATTTGCTGAACCGCGCGCAT 205
 |||||
 1582ValAlaSerValLySProProSerAsnSerSe 1592
 204 AAC.....ACACGCGCGGATCTTT..... 183
 |||||
 1592 rSerValGlylnAsnIleGluValThrValProSerTyrlPheProValG 1609
 182TTGTCTCAACAGCATGGGCTTTTTCAGCGCATG... 144
 |||||
 1609 luGlnSerLySerValThrGlnAlaProIleProProAlaSerSerPro 1625
 143CTTCTTGAAGTTCATGAGGGCGCATACCGGCATATTCTTC 100
 |||||
 1626 SerAlaProSerlleProSerlleProSerThrleuProValAlaProSe 1642
 99 GCCAAGCAGCGCGCATTCGTAATGCGCGGCGGTG 63
 |||||
 1642 rSerSerSerValProSerSerProSerAlaProSer 1654
 seq_name: sp_invertebrate:P91495

seq_documentation block:
 ID_P91495 PRELIMINARY; FRT: 1847 AA.
 AC P91495;
 DT 01-MAY-1997 (TREMblrel. 03, Created)
 DT 01-MAY-1997 (TREMblrel. 03, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE HYPOTHEICAL 199.6 KDA PROTEIN.
 GN T23H2.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peioderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP STRAIN-BRISTOL N2;
 RC MEDLINE-99069613; PubMed-9851916;
 RX None;
 RA "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.",
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;

RA Wamsley P., Bradshaw H.;
 RT "The sequence of C. elegans cosmid T23H2.";
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RT "Direct Submission.";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; 080033; AAC48199.1;
 DR InterPro; IPR003343; B19_2.
 DR Pfam; PF02368; B19_2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 1847 AA; 199638 MW; 1A5BDAD1261D549A CRC64;

alignment_scores:
 Quality: 110.00 Length: 511
 Ratio: 0.464 Gaps: 26
 Percent Similarity: 46.380 Percent Identity: 22.114

alignment_block:
 US-09-303-518D-125 x P91495 ..

Align seg 1/1 to: P91495 from: 1 to: 1847

4 ATTAATCAAAAGGCTTAACCTGCCCATCGCGGCGAGACGAGCA 53
 |||
 818 lIeThVallylsylslyleuAsnVal.....GluProAlaR 830
 54 AGCGTTTACGACGCGCGCGCATTTACGAGTCCGTTCTT..... 96
 |||
 830 glEuValleuTrpAsnGluAlaValSerLySgLyThrPheThrIleThrg 847
 97GGCGAAGATATGCCGATGCGCCCTCGATGAAGTCAAG 138
 |||||
 847 lyGlySerGlyHisPheHisValAspAsnLeuProThrSerAspSerPro 863
 139 GAAGCGCATCGCGTCAAAAGGCGCAAGTGTCTTTGAAGACAAAGAA 188
 |||
 864 ValAlaIleAlaLeuArgAlaArgSerLeuThrValThrProLyAsnAs 880
 189 TCCGCGGCGGTGTACT.....G 208
 |||
 880 nglyGlnValAsnLeuArgIleSerAspAlaCysLeuValGlylnHisA 897
 209 GCGCGCTTCAGCGCAAAATCGCGCGATTCACCGTGGCAAAAGCGGTA 258
 |||
 897 laAspAlaSerValArgIleAlaAspIleHis..... 907
 259 CTTGAGTCAGTCGTGAT.....GCCGTTGAAGCGACGACGA 296
 |||||
 908SerleuAlaIleAspAlaProGlnPheValGlnIleGlylnG 922
 297 AATGAGTTTGAACGCTACGACCTGAAGCGGTGCAAACTTAAGCGGCG 346
 |||||
 922 uValGluValGluIleuAlaGlnAspGlu...ThnglyAlaSerPheg 938
 347 AAGAGTCCCGCGCAACCTGATCCATCCGCTTGTGGATGGCGTGGCG 396
 |||||
 938 luLySglnHisArgProLeuAlaAspAlaClnLeuAspAlaSerAsn 954
 397 ACCGCTCGCTCAAAATTCCTGCGCGATGCCGAGCGGTTGCGCAT 446
 |||||
 955 HisValIleleuThrLyS.....ValAspLyLeuArgTyrlle 968
 447 CTTGTCATGCGATGACACCACTCCGCTGCGTGGCGACCGCTACAGGTC 496
 |||||
 968 uArgAlaAsnSerIleGlyThrValSerLeuSerAla.SerSerLyS 984
 497 TTATCAAGAACCGCGCGGATTTCAACCGCGCGCTGTGATTGAGC 546
 |||||
 985 SerSer.....GlyArgValLeuSerSerArg..... 993

```

547 CGTTTGACCGAAGCAAAATCCATGTTGTAAGCAGCTGGCGAGACT 596
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
994 .....ProH1sthrValGlnIlePheSerProIlePheLeuGlnProL 1008
597 GCCGCTGTAATGCTGCCAACATCGAAGAACATGAA.....TTGGCG 639
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1008 ysArgLeuThrLeuIleProAspSerLysPheGlnLeuGlnValValGly 1024
640 GCCCGCATCTGCGC.....GTTTGAAGTGGCAGCAGCAT 674
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1025 G1yProGlnIleProIleProIleProLeuAspPheSerLeuAsnSerMetI 1041
675 TCATTTCATCGAGCCG.....GTGGCGCGCA 700
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1041 eAlaSerIleGlnProAsnAlaLeuIleThrSerSerGlnLeuGlyTyrT 1058
701 ATAAACGCTGTGGACCATCAATTATCAAGAT..... 732
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1058 hAlaIleThrGlyThrValArgValGlyAspGlnHisValThrLeuAsp 1074
733 .....GTAATTACCATGGCCGTTTGTTCACACA..... 762
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1075 ThrValValLeuArgValAlaSerLeuGlyGlyIleIleLeuSerAlaSe 1091
763 .....GGCGCTGTGAAC..... 774
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1091 rSerArgLysValGlnIleThrGlyArgValAlaAsnLeuArgLeuArgGly 1108
775 .....ACGAGCGCGTGTATGCGCTGATGCTGTCT... 804
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1108 alIleAlaGlyAlaGlnAspGlnIleProPheAlaPheGlyGlyAlaIle 1124
805 .....CAAGTCACAAACCGCGCTTGGCGTAC 833
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1125 TyrProPheLysValThrTrpSerValSerAspProSerValLeuPheH 1141
834 CGTT.....TTGGTGGCAAGTATTCGCAAAATTAATCGCGCGCAATTGG 877
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1141 rLThrHisProLeuGlyGlyAspValGlnIleProThrAspAsnGlnPheA 1158
878 TTGACACAGACAAACCGCGTATTCGCGTTCGTTAATGACCGCGCAT 927
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1158 lAlleTrpPheAsnAlaIleArgGlySerValThrValLysAlaVal.. 1174
928 ACACA.....GGCGGCGACGATTAATTGGACGCTACAC.. 963
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1175 ValGlnLeuAsnGlnLysAlaArgLysHisPheThrGlyArgThrSerTh 1191
963 ..... 963
1191 rPheThrAlaGlnIleThrIleThrValGlnAspGlyLeuSerLeuValG 1208
964 .....AATCAGATTCCTGTTATCGAAGAGCGCGACG 996
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1208 lArProGlnLeuAspIleAsnThrValArgValAlaIleProAsnSerGlnLeu 1224
997 AAAGAGCTGTGGCTGGGTTGGCGCGACG.....CCGACACAAATA 1037
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1225 LysMetValThrAlaTrpSerGlnAlaSerPheSerValProSerAspPh 1241
1038 CTCC.....ATCAGCGGTACAAACCTCGGCATTCCTGTAAGAACAAAC 1081
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1241 eSerSerArgIleValIleSerAlaAspGlyHisLeuIleThrAsnGly.. 1257
1082 TCTTCAGTTCAACACAGCC.....GTCAACGCGGCG 1113
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1258 .....LysGlnGlySerAlaAlaIleThrValArgAsnValAlaSerPro 1272
1114 GACGCG.....GCCATGTCGCGATGTGTTACTTACGAGCGCGGATGCC 1157
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1273 AspAsnGlnIleThrValLeuIleProVal.....ThrValSerArgValAlaLase 1288

```

```

1158 CTTGGATATCTGCCACCGCTGTTGGCG 1188
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1288 rLeuAspValHisProThrIleGlnLeuLys 1298

```

seq_name: sp_bacteria:Q92A63

seq_documentation_block:
ID Q92A63 PRELIMINARY; PRT; 270 AA.

AC Q92A63;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
DE HYPOHETICAL 30.2 KDA PROTEIN.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B1940;
RX MEDLINE=99005251; PubMed=9790590;
RA Claus H., Frosch M., Vogel U.;
RT "Identification of a hot spot for transformation of Neisseria meningitidis by shuttle mutagenesis using signature-tagged transposons."
RT Mol. Gen. Genet. 259:363-371(1998).
RL RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=B1940;
RA Claus H.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF000250; CA003966.1; --
KW Hypothetical protein.
SQ SEQUENCE 270 AA; 30192 MW; 612DE74D02BA09AA CRC64;

alignment_scores:

Quality: 109.50 Length: 382
Ratio: 0.720 Gaps: 25
Percent Similarity: 39.791 Percent Identity: 26.178

alignment_block:

US-09-303-518d-125 x Q92A63 ..

Align seg 1/1 to: Q92A63 from: 1 to: 270

```

32 CCATCGCGGCGAGACCGAAGCCGTTTACG..... 64
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6 ProTyAlaSerAspGlyLysLeuProAspThrHisGlnGlnIleProcy 22
65 .....ACGCGCGCGCATTAACCGAAGTCGCGTGTCT 95
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
22 sProProValSerAlaProThrAlaLysProValSerGlySerLys.... 37
96 TGGCGAAGATATGCGGTATGCGCCCTCGATGAAAGTCAAGGAGCG 145
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
38 .....LysProAsnSerMetSerProLysAla 46
146 ATGCGGTCA...AAAAGCCCAAGTGTGTTGAAGACAAAGAAATCCG 192
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
47 SerSerSerAlaLysAsnAlaLysGlyCysLeuLysProLysThrIle.. 62
193 GCGGTGC.....TGTTCCTGCGCGCGGTTC 218
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
63 .....TrpGlnAlaArgLysAsnProLysThrIleCysProArgLeuP 78
219 AGGCAAAATCGCGCGGATTCACGTGCGGAAAGCGGCTACTTCAGTCAG 268
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
78 hArgMetSerAsnSerPheThrValSerAlaArgAlaProSerAlaArg 94
269 TCGTGATGTCGCGTGAAGGCAAGCAAGAAATCGATTGACCGTACGCA 318
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
95 AsnArgPheProVal.....ThrLysSerProAlaSerSerThrAl 108

```



```

163 LeuMetAlaGluLysLeuGluGluIlePheSerGlyPheGlnIleLeu 179
520 .....TTCAACGCGCGCGTGTGATTTAGCCGTTTGACG 556
179 PylsIlePheLysPheLysGluIleIleIleIleSerAsnLysAsp 196
557 AA.....CGCAAAATCCATGTTGTAAAGCAGCTGGCGCA 591
196 yslLeuLysLysGluPheGluLysLeuSerIlePheLysAsnAlaGlySle 212
592 GAGCGCGCGCT...GAAATGCTGCCACATGAAACACATGATTCGG 638
213 LysIleLysSerLeuGluAsnAlaTyProTyThrAsnHisGluMet... 228
639 CGGCGCGCATCTGCGCGTTGAGTGCAACGCAATTCATTCATC... 684
229 .....IleMetHisPheLeuTyR... 735
685 .....GAGCGCGTGGCGCGCAATTAACCGTGTGGACG 717
235 snAsnLysAsnThrLysAspAspIleAsnProHisAsnHisLeuLeu 251
718 ATCAATTATCAAGATGTAATTACCATTTGCCGTTGTTGCAACAGCGCG 767
252 AlaAsnIleGluAspLeuTyAsnAlaAsnLeuValIleLysAsnAsnAs 268
768 TCTGAACACCGACGCGCGTATTGCCCTAGTGTTCTCAAGTCACACAAAC 817
268 nProTyLysGluLysPheValAlaIleAsnGlyAsnLysLysIleLys 285
818 CGGCGCTCTTGCGTACCGTTTGGTGGCGAAAGTATCCAAATTAATCTCG 867
285 errArgIleLeuLysValIleLysIleGlyThrSerPheSerGlnLeu...Ile 300
868 GCGGAATTGGTTGACACA 885
301 AsnGluLysIleAspPhe 306

```

seq_name: sp_invertebrate:Q22579

seq_documentation_block:

ID Q22579 PRELIMINARY; PRT; 1844 AA.

DT 01-NOV-1996 (TEMBLrel. 01, Created)

DT 01-OCT-2001 (TEMBLrel. 18, Last sequence update)

DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)

DE HYPOHETICAL 187.1 KDA PROTEIN.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditidae;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA MEDLINE=99069613; PubMed=9851916;

RA None;

RT "Genome sequence of the nematode C. elegans: a platform for

investigating biology. The C. elegans Sequencing Consortium.,"

Science 282:2012-2018(1998).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Favello A.;

RT "The sequence of C. elegans cosmid T19D12.,"

Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Waterston R.;

RT "Direct Submission.,"

Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

EMBL: U41263; AAC24428.2; .

KW Hypothetical protein.

SO SEQUENCE 1844 AA; 18713 MW; 6151A72777517P5 CRC64;

alignment_scores:

Quality: 109.50

Ratio: 0.788

Percent Similarity: 43.849

Percent Identity: 24.606

alignment_block:

US-09-303-518D-125/rev x Q22579

Align seg 1/1 to: Q22579 from: 1 to: 1844

935 CTTGTGTAATTCGCGCGTTCATATACCAACCGGAATACGCGGTTC 886

310 ProValValThrThrSerThrSerThrGlnGlyIleSerThrThrAl 326

885 TGTCACCAATTCGCGCGAGTAAT..... 858

326 aGlnAlaThrProSerSerValIleProThrThrGlnThrGln 343

857TGCAATCTTTCGACCAACCAACGTCAGCAAGAGCGCGT 816

343 InArgProThrSerThrGlyIleProSerThrValSerThrSerGln 359

815 TTGTGACTTGAACACCACTAGGCAATACGCGCGTGTTCAGACG 766

360 ThrSerSerThrSerProIleProSerThrGlnThrSerSerAl 376

765 GCCTGTGCAACCAACGCGCAATGATTAATCTTGATTAATGATG 716

376 aPro.....SerT 379

715 TCACACGCTTTATTCGCGCGACCGCTGATGAATGATGCTG 666

379 hrTyThrSerAsnThrProSerProThrThrLeuLeuThrSer 395

665 CCACTCMAACCGGACGAGATGCGCGCGCAATTCATGTTGCATGT 616

396 ThrIleAlaProSerThrGlnGlyValProThrSerSerSer 410

615 GGCACATTTTCAGACGCGACGCTGCGCGACGCTTCAACACATGA 566

411SerSerProAsnSerThrProThr.... 418

565 TTTGCGTTCGTCMAACGCGCTCAATACCAACAGCGCGCTTGAATCC 516

419ThrThrIleThrProGlyAlaProSer 427

515 TCGCGCGCTTCTTGATTAATGACCGTATGCGTACGACGCGATG 466

428 Ser.....ThrLeuGlySerSerSerThrIleVal 438

465 GTCCATGCAATTCAGACGAGATGCGCAACGCGTGCATGAGCGGAA 416

438 lserThrThrIleThrPro...SerThrProLysValSerThrLeuThrL 454

415 TTTTGTGACGAGCGGCGTGCAGCGCAAGTCACCAACCGATGATC 366

454 eu.....SerGlnSerProThrPro..... 460

365 AGCTTGGCGGCACTTCTTCGCGCTTAAGTTGCGAGCGTTGAGTGC 316

461ThrSerThrProLeuValValSerSerSer..... 471

315 GTAGCGTTCAAACTTCATTTGCTGCTTCAACGCGCAATCAGACTG 266

472SerGlySerSerSerThrValValThrSerThrIleThrPro. 485

265 ACTGAATGACGCGCTTTCGACGCGTGAATGCGGCGCATTTTGCCTGAA 216

486SerThrGlnGly.....ValProThr 492

[illegible]

491 GTAGGGTCGGCAGCCAGCGGATGTCTCATTCGCATTGACGAAGATGCC 44
|||::|||:::||||||| ||:::
1668 ValAlaserSerThrThgIyeuValSer.....ThrSerTrva 166

```

441 GAACGGCTGGCATCGACGCGAGCATTTTGTGAACGAGCGGTGGCA 392
    : ::::::::::::::| | | | |
1681 lProserSerThrGlyThrMetGlySerThrSerSerGlyThrValGly 1698
    : ::::::::::::::| | | | |
391 GCGCAGTCCACCAACCGGATTGGATC.....AGTTGGCGCGCAGCTTCT 348
    : ::::::::::::::| | | | |
1698 erThrIleSerGluSerSerThrAlaIleSerAlaSerSerGlnThrGly 1714
    : ::::::::::::::| | | | |
347 TCGCGCGTTAAGTT.....GCCAGCGCTTCAGTGGCTAGCGTCAAA 304
    : ::::::::::::::| | | | |
1715 SerThrValThrMetGlySerSerSerThrSerGlyValSerThrSer 1731
    : ::::::::::::::| | | | |
303 CTGATTGTCGCTGGCT.....TCACGGCGCA 275
    : ::::::::::::::| | | | |
1731 rAlaSerSerGlyGlnProGlnMetSerThrSerGlnGlySerSerAlaG 1748
    : ::::::::::::::| | | | |
274 TCACGACTGCTGACGCGCTTTGCCACGCGTGAATCGCGCGAT 225
    : ::::::::::::::| | | | |
1748 lSerThrValValSerSerThrAlaSerProAlaIleSerSerThrAla 1764
    : ::::::::::::::| | | | |
224 TTGCTGAGACCGCGCGCAGTAAACACACGCGC.....CGATT 187
    : ::::::::::::::| | | | |
1765 ProSerSerThrGlyThrMetSerSerThrSerSerGlyThrValGly 1781
    : ::::::::::::::| | | | |
186 CTTTGTGCTTCAACAGCATTGGCTTTTTCAGCGCATCGCCTTCT 137
    : ::::::::::::::| | | | |
1781 rThrMetSerGlnSerSerThrAlaIleSerThrSerThrSerThrGly 1798
    : ::::::::::::::| | | | |
136 TGACTTTCATCGAGGCGCATACCGCATATTTTCGCAAGCAGCGCG 87
    : ::::::::::::::| | | | |
1798 erThrValThrLeuGly.....SerSerSerThrSerSerAlaGln 1811
    : ::::::::::::::| | | | |
86 ACTTCGGTATGCGCGCGCGCGTGG 63
    : ::::::::::::::| | | | |
1812 MetSerThrSerGlnGlySerSer 1819
seq_name: sp_Bacteria:Q9KXT0

seq_documentation_block:
ID Q9KXT0 PRELIMINARY; PRT; 960 AA.
AC Q9KXT0;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-OCT-2001 (TEMBLrel. 18, Last annotation update)
DE PUTATIVE SUGAR HYDROLASE.
GN SCSH4_09C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_taxid=1902;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-A3(2);
RA Murphy L., Harris D.;
RA Submitted (May-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RA Submitted (May-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN-A3(2);
RA MEDLINE-97000351; PubMed-8843436;
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinsahl H., Hopwood D.A.;
RA "A set of ordered cosmids and a detailed genetic and physical map for
RA the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RA MBL. Microbiol. 21:77-96(1996).
DR EMBL; AL35513; CAB91121.1;
DR InterPro; IPR001764; Glyco_hydro_3.
DR InterPro; IPR002772; Glyco_hydro_3C.

```

```

DR Pfam; PF00933; Glyco_hydro_3; 1.
DR Pfam; PF01915; Glyco_hydro_3_c; 1.
DR PRINTS; PR00133; GLHIDRLASE3.
KW Hydrolase.
SQ SEQUENCE 960 AA; 102056 MW; 5332C02428BF39A2 CRC64;

```

```

alignment_scores:
    Quality: 109.00      Length: 470
    Ratio: 0.522        Gaps: 27
    Percent Similarity: 44.468      Percent Identity: 23.191
alignment_block:
US-09-303-518D-125 x Q9KXT0

```

Align seg 1/1 to: Q9KXT0 from: 1 to: 960

```

46 CCGGAGCAGACCGCTTACGAGCGCGCGCCATTCAGGAAGTCGGTCT 95
    : ::::::::::::::| | | | |
319 ProGluAsnAspProTyrSerGlyAlaIleAspThr..... 330
    : ::::::::::::::| | | | |
96 TGCGAAGATATAGCGGTATCGCGCCCTCGATGAATCAAGCAGAGCG 145
    : ::::::::::::::| | | | |
331 .GlyAlaAspThrAsp.....ThrProGluHisArgAlaLeuAlaArg 345
    : ::::::::::::::| | | | |
146 ATGCGGTCAAAAAGAGCGCAAGTCTGTTGAAGCAAAAGATCGGGC 195
    : ::::::::::::::| | | | |
345 sPAlaIleGluGlnAlaValValLeu.....LysAsnAspGly 358
    : ::::::::::::::| | | | |
196 GTGGTGTACTACGCGCGCTTCAGGCAAAATCGCGCGATTCCGCTGG 245
    : ::::::::::::::| | | | |
359 ValLeuProLeuAlaPro..... 364
    : ::::::::::::::| | | | |
246 CGAAAAGCGCGTACTTCAGTCAGTCGTATGCGCTGAAGGCAAGAGC 295
    : ::::::::::::::| | | | |
365 .GluThrArgValAlaValAlaGlyLeuLeuAla.....AspG 377
    : ::::::::::::::| | | | |
296 AATAGAGTTTGACGCTACGACCTGAGAGCGCTGCAAACTTAAGCGC 345
    : ::::::::::::::| | | | |
377 LuCysLeuAspTrpTyr.....SerGly 385
    : ::::::::::::::| | | | |
346 GAAGAGAGCGCGCAACCTGATCCAAATCCGGTTTGG..... 384
    : ::::::::::::::| | | | |
386 ThrLeuIleHisArgSerThrProLeuGlnGlyLeuTyrGluArgPheG 402
    : ::::::::::::::| | | | |
385 .....ACTGCGCTCGCA 397
    : ::::::::::::::| | | | |
402 yAlaAspArgValSerPheAlaGluGlyValAspArgValArgLeuArg 419
    : ::::::::::::::| | | | |
398 CC.....CGTCGTTGAGCAAAATTCCTGCGCGATGCG..... 432
    : ::::::::::::::| | | | |
419 hrAlaAspGlyArgPheLeuHisValLeuProAlaAspSerAlaSerAla 435
    : ::::::::::::::| | | | |
433 .....GAGCGCTGCGCATCTTCGTCA 455
    : ::::::::::::::| | | | |
436 GluAlaProGlyThrGluGlyAlaLeuAspPro...AlaLeuLeuAla 451
    : ::::::::::::::| | | | |
456 TGCATGACACCAATCGCTGGTCGCGACCTACGATCATTAATCAAG 505
    : ::::::::::::::| | | | |
451 yArgThrAspLeuProProLeuThrThrAsp...AlaValGlyThrGlu 467
    : ::::::::::::::| | | | |
506 AAGCGCGGAGATTTCAAACGCGCGCTGTGATTAGCGCTTGGACC 555
    : ::::::::::::::| | | | |
467 euAlaLeuIleAspTrpGlyGluGlyValLeuThrLeuArgAlaSerAsp 483
    : ::::::::::::::| | | | |
556 GAACGCAAAATTCATGTTGTAG.....GCAGCTGGCGCAGA 593
    : ::::::::::::::| | | | |
484 GlYArgTyrLeuSerValAlaGluAspGlyPheValArgAlaSerAlaAs 500
    : ::::::::::::::| | | | |
594 CGTGGCGTCT.....GAAATGCTGCCAATCGAAGACAGCATG 631
    : ::::::::::::::| | | | |
500 pGlnProGlyGlyTrpValValGlnGlnThrPheArgLeuGlnProHisA 517
    : ::::::::::::::| | | | |

```

```

632 AATTGGGGGGGGGGGATCTCCGGTTTGAATGCGACGACATTCATTTG 681
    |||
517 SPASp61Y.....HisLeu 521
682 ATCGAGCGCGTGGCGCGAATAAACCGTGTGACG..... 717
    |||
522 LeuATGHisLThrGlyThrGlyArgProValGlnValAlaAlaAspGlyVa 538
    |||
718 ...ATCAATTATCAAGATGTAATACATTTGGCGCTTTGTTTGCACAG 763
    |||
538 LysValAlaAlaProAspAlaAlaGlyAlaAlaAlaGlyGlyVal 555
    |||
764 GCGGTCGTAACACGCGCGGTGATTCGCTAGTGGTTCACAGTCAAC 813
    |||
555 LysSerAlaAspAlaGlu...ValPheGlnLeuValValGluArgGly 570
    |||
814 AACCGCGCGCTCTTGGCGTACGCTTTGGGTGGGAATGCAATTCAC 863
    |||
571 GluAspAlaValThrArgValAlaAlaGlyAlaAspAlaValValVa 587
    |||
864 TGGGGCGAATGCTTGACACAGACACGCGGTGATTCGCGTTCGAT 913
    |||
587 LalaGly.....AsnAspProHisI 594
    |||
914 TGACGCGCGGATTCACACAGCGCGCGCATTTATTGGAGCGTACACAG 963
    |||
594 LeAsnGlyArgGluThr.Glu.....AspArgThrTh 604
    |||
964 AATCAGATTCGCTGATTCGAGAGAGCGCGCAAGAGCTGTGGCGTG 1013
    |||
604 rLeuAlaGlnProAlaGlnGlnLeuArgLeu..... 614
    |||
1014 GATTGCGCGCGACCGGACAAATATCATTCACGCGTA..... 1051
    |||
615 ..LeuArgAlaAlaArgAlaAlaAsnProAlaThrValLeuAlaLeuVal 630
    |||
1052 .....CAACCTCGCGCATTTCTCTGAAAAAACAATCTTCAAGTTCAAC 1095
    |||
631 SerAlaTyPProTyAlaVal.....GlyValGlnTh 641
    |||
1096 ACAGCGCTCAACGCGCGCGCGCGCATGTCGCGATTGGTACATTACA 1145
    |||
641 rLeuProAlaValLeuThrPThrAla.HisGly..... 651
    |||
1146 GCGCGTGATGCGCTTGATTCCTGCCACCGCTTTGGCGGATTTAA 1195
    |||
652 .....GlyGlnAlaAlaGlyThrAlaLeuAlaArg.Ileu 663
    |||
1196 TCGTCGCGGATACGACGCGCGCGCATTTG...GTTGCTTGAATTG 1242
    |||
663 euAlaGlyAspValSerProAlaGlyArgLeuProGlnThrTrpTyrSer 679
    |||
1243 GACGAGAAAGACCTGCTTTGTGCGCTTGTCTGCCCGGCAATACGA 1292
    |||
680 AspAspAlaAspLeu.....ProGlyLeuLeuAs 689
    |||
1293 ATAC 1296
    |||
689 pTyr 690
    |||

```

seq_name: sp_human:09UHA8

seq_documentation_block: PRELIMINARY; PRT; 2296 AA.

AC 09UHA8;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE SPLICING COACTIVATOR SUBUNIT SRM300.
 GN SRM300.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20132238; PubMed=10668804;
RA Blencowe B.J., Baurer G., Eldridge A.G., Issner R., Nickerson J.A.,
RA Rosolina E., Sharp P.A.;
RT "The Srm160/300 splicing coactivator subunits."
RL RNA 6:111-120(2000).
DR EMBL; AF201422; AAF21439.1;
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PRO1217; PRICHEXTENS.
SQ SEQUENCE 2296 AA; 251964 MW; 17C0BD4FA10A9CF9 CRC64;

```

alignment_scores:
 Quality: 111.50 Length: 435
 Ratio: 0.578 Gaps: 20
 Percent similarity: 44.368 Percent identity: 23.678

alignment_block:
 US-09-303-518D-125 x 09UHA8 ..

Align seg 1/1 to: 09UHA8 from: 1 to: 2296

```

18 AGGCTAAACCTGCCATCGCGCGCGACGACGCGCTTACGACG 67
    |||
483 ArgSerArgSerProAlaThrAlaLysArgGlyArgSerArgSerArgTh 499
    |||
68 GCCC.....GGCATTACGAGATCGC...CTTCTTGGCGAGAA 105
    |||
499 rProThrLysArgGlyHisSerArgSerArgSerProGlnThrArgGly 515
    |||
106 TATGCCGCTATGCGGCC.....CTCGATGAAGTCAGAGAGCGGA 146
    |||
516 ..SerArgSerAlaGlnArgTrpGlyArgSerArgSerProGlnThrArg 531
    |||
147 TGCCTGCAAAAAAGCCAGT..... 167
    |||
532 GlyArgSerArgSerProGlnArgProGlyTyrPserArgSerArgAsnTh 548
    |||
168 .....GCTGTTGAAGACAAAAGAACATCCGGCG 195
    |||
548 rGlnAlaArgAlaGlyArgSerArgSerAlaArgGlyArgSerHisSer 565
    |||
196 GTGGTGTTTACTCGCGCGCTTCAGCAAAATCGCGCGATTCACCGTG 245
    |||
565 rGserProAlaThrArgGly.....ArgSerArgSerArgThrProAla 579
    |||
246 CGAAAGCGCGTACT.....TCAGTCAG 268
    |||
269 TCGTGAATTCGCGTTGAAGCGACGACGAAATCGATTTCAGCGTACGA 318
    |||
580 ArgArgGlyLysArgSerArgSerArgThrProAlaArgArgArgSer 596
    |||
319 CCTAAGCGCTGCGCAACTTAAGCGCGGAAGAGTGGCGCGCAACTGTAT 368
    |||
613 LysArgSerArgSerArgThrProAlaArgArgArgSerArgThrArgSer 629
    |||
369 CCAATCGGTTTGTGAGTGGCGTGGCGCACCGCTC...CGTTACGACAA 414
    |||
630 ProValArgArgArgSerArgSerArgSerProAlaArgArgSerArg 464
    |||
415 ATTCCTGCGCTGATGCGCGAGCGCTTGGCATCTTGTCAATGCGATGA 464
    |||
646 gSerArgSerArgThrProAlaArgAlaGlyArgSerArgSerArg...T 662
    |||
465 CACCAATCGCTGCGTGGCGCGACCTTCAGTCATATTAAGAAAGCCGCG 514
    |||
662 hrProAlaArg.....ArgGlyArgSerArgSerArgThrProAla 675
    |||
515 AGGATTTCAAAAGCGCGCTTGTGATTTGACCGCTTGCAGCAAGCGAA 564

```



```

676 ArgArgSerGly.Arg.....SerArgSerArgT 665
565 ATCCATGTTTGTAGCGACGCTGGCGGACAGCTCCCTCGAATAATCTCG 614
665 hrpP01A1ArgArgGlyArgSerArgSerArgThrProArgArgGlyArg 701
615 CAACATCGAACAACATGATTCGGCGCGCCGATCCCTCGCGGTTTACGG 664
702 SerArgSerArgSerLeuValArgArgGlyArgSer..... 713
665 GCAGCGACATTCATTCATTCGACCGGT.....C 693
714 .HisSerArgThrProGlnArgArgGlyArgSerGlySerSerSerGluA 730
694 GCGCGCAATAAACCGTGTGACATTCATTCAGATCAACAGATTAACCT 743
730 rglYsAnLysSerArgThrSerGlnArgArgSerArgSerSerSer 746
744 TGCGCGCTTTGTCACACAGCGCGCTGAACACGACGCGCTGAT..... 788
747 ProGluMetLysSerArgThrSerSerArgArgSerArgSerLeuSe 763
788 ..... 788
763 rSerProArgSerLysAlaLysSerArgLeuSerLeuArgArgSerLeu 780
789 .....TGCCCTAGGTGGTCTCTCAAGTCAACAAACCGCGCT 824
780 enGlySerSerProCysProLys.....GlnLysSerGlnThrPro 794
825 CTTCGCTACCGTTTGGGTGCAAGATATCGCAATTACTCGCGGCGAAT 874
794 ..... 794
875 TGCTGACACAGAACCGCGTGTATTCGGTTCGTAATGACGGCGG 924
794 ..... 794
925 ATTACACAGCGCGCAGATTATTGGACGCTACACAAATCAGATTTC 974
795 .....ArgArgSerArgSerGlySerSerGlnProLysAlaLysSer 808
975 CGTTATCGAAGAGCGCGCAGCAAGACCTGTGCGCGGTTCGGCGG 1024
809 ArgThrProProArg.ArgSer.ArgSerSerSerProProProLys 824
1025 AGCCGCAACAAATCTCCATCGACGCTACACCTCGGCCATTTCGTAAA 1074
825 GlnLysSerLysThrProSerArg.....G1 833
1075 AACAACTCTTCAGTTCAACACAGCGGTCAACGCGCGGCGACCGCCCA 1123
833 nSerLysSerSerSerProHisProLysValLysSerGlyThrPro 849
seq_name: sp_bacteria:Q93RW3
seq_documentation_block:
ID Q93RW3 PRELIMINARY: PRT: 406 AA.
AC Q93RW3;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE HYPOHETICAL 44.7 KDA PROTEIN.
GN SC139.10C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Warren T., Harris D.;

```

```

RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=9700351; PubMed=8843436;
RA Redenbach M., Kiser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinsht H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL MOL. Microbiol. 21:77-96(1996).
DR EMBL; AL591322; CAC38798.1;
KW Hypothetical protein.
SQ SEQUENCE 406 AA; 44724 MW; E64CFBBDCTA632F CRC64;

```

```

alignment_scores:
Quality: 108.00 Length: 410
Ratio: 0.632 Gaps: 21
Percent Similarity: 41.707 Percent Identity: 22.683

```

alignment_block:

US-09-303-518d-125 x Q93RW3 ..

Align seg 1/1 to: Q93RW3 from: 1 to: 406

```

102 AGAATATGCCGGTATGCGCCCTCGATGAAGTCAAGAGCGCATCGC 151
111:.....:|||||.....:|||||
33 ArgValAlGlyLeuHisAlaProLeuLysPgl...ArgGlyThrGlnCysAr 48
152 TCAAAAAAGC..... 161
48 glnAlaAlaLeuHisGlyProAlaArgValArgThrValValAla 65
162 ..CCAAAGTCTGTTGAAGACAAAAGATCGGCGCTGTGTTTACTGC 209
65 sPrProHisLeuAlaArgGlnArgAlaAlaAlaArg..... 77
210 GCCGCTTCAGGCAAAATCGCCGC.....GATTCACCGCTG 244
78 .....ArgGlnArgGlnArgValAlGlyLeuGlnGlyThrGln 91
245 GCGAAACCGCGTACTTACGACTGCTGATTCGCTTGAAGCAAGCAG 294
91 sGlnArgGlyArgValAlaArgValAlaArgValAlaGlnCysArgHis 108
295 GAATCGAGTTTGAAGCGTACGCAAGCTGAGCGCTGGCAAACTTAAGCG 344
108 rGlnArgValAlaProArgLeuValHisValAlaArgLeuGlnAlaAlaGlnArg 124
345 CGAAGAGTGGCGCGCAACCTGATCCAAATCGGTTGTGAGTGCCTGC 394
125 GlnArgAlaGlyPro..... 129
395 GCACCGCGCGCTTACGCAAAATTCCTGCGCTGATCCGACCGCTTGC 444
130 .....GlyGlyArgArgArgGlyArgValAlaProArgH 140
445 ATCTTCGTCAATGAGTACACCAATTCGCGTGGCGG..... 485
140 is.....HisArgAlaGlyAlaLeuArgLeuAspAla 150
486 .....CCCTACGCTATTATCAAGAGCGCGGAGATT 520
151 LeuAspAlaGlyLeuGlnArgGlnHisLeuGlnLeuArgArgValAl 167
521 TCAACGCGCGCTGTGTTGATTGAG.....CCGTTTGACCGAA 558
167 nglnArgLeuAlaProAlaValAlaProAlaLeuLeuProValAlaGlnGlu 184

```

```

RC      TISSUE=LIVER;
RA      Kristie T.M., Dashner R.;
RT      "cDNAs encoding the mouse homolog of the human transcription factor C1
KT      (HCF).";
RL      submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.
DR      EMBL; U53925; AAB0163.1; -.
DR      MGD; MG1:105942; Hcfl.
DR      InterPro: IPR003961; FN_III.
DR      InterPro: IPR01798; Kelch.
DR      Pfam; PF00041; fn3; 2.
DR      Pfam; PF01344; Kelch; 4.
DR      SMART; SM00060; FN3; 1.
SQ      SEQUENCE      2045 AA; 210537 MW; B89CC2FDA35B969F CRC64;

alignment_scores:
      Quality: 109.50      Length: 617
      Ratio: 0.423      Gaps: 25
      Percent Similarity: 41.977      Percent Identity: 19.935

alignment_block:
US-09-303-518D-125/rev x Q61191 ..

Align seg 1/1 to: Q61191 from: 1 to: 2045

1343 CAGCCTTCCTTCATCATGGTTCCAGCAGCACTTGGCGAAGAGGGGCGGTA 1294
      ::::::::::::::::::::: :::::::::::::::::::::
747 LysProThrHisLeuGlyLeuSerVal.....SerProse 759
1293 TTGCGATTGGCCCGGAGAGAGCAAGCACTGCACAAAGCAGGCTCTTCGT 1244
      ::::::::::::::::::::: :::::::::::::::::::::
759 rtrtrhrlysproglythrthrthrllelelythrleprometsara 776
1243 CCATTCACGACACCAATGCTCGCGCGGTGCGTATGCGCGACGATT 1194
      ::::::::::::::::::::: :::::::::::::::::::::
776 lallelethrghnalaglyalatrnglyalthrSerSerProglylle 792
1193 AAATCG..... 1188
      :::::::::::::::::::::
793 LysSerProlethrlethrlethrthrlyrlyValmettnserglyThrGl 809
1187 .....CGAAAGAGGCGG93AGCA 1166
      :::::::::::::::::::::
809 yalaProalaLySlelleThrAlaValProLySleAlatrnglyHisG 826
1165 TATCCAGGCGCATCAGCGC..... 1146
      :::::::::::::::::::::
826 LyGlnGlnGlyValThrGlnAlaValLeuLySglYAlaProGlyInPro 842
1145 .....TCGTAAGTACCAATCGCACACATGCGCGGTGCCGCC 1108
      :::::::::::::::::::::
843 G1YThrLleLeuAArgThrValPrometGlyGlyValArgLeuValThrPr 859
1107 GTTGAGC.....GCTGTGTGAAGTTCAGAGAGTTGT 1076
      :::::::::::::::::::::
859 ovalThrValSerAlaValAlaLySProAlaValThrThrleu..... 872
1075 TTTTCAGGAATGCGCGAGGTTGTAGC..... 1047
      :::::::::::::::::::::
873 .....ValVallySglYThrThrGlyValThrThr 882
1046 .....GTGATGGAGTATTTGTGCGGTGCGGCGC 1018
      :::::::::::::::::::::
883 LeuGlyThrValThrGlyThrValSerThrSerLeuAlaGlyAlaGlyAl 899
1017 AACCCAGCCGACAGCTTTGTGCGGCGCTTCGATAC.GCAATC 969
      :::::::::::::::::::::
899 ahISerThrSerAlaSerLeuAlatrhrProlethrThrLeuGlyThrI 916
968 TGATTTGGTAGCGTCCCAATATC..... 943
      :::::::::::::::::::::
916 lealatrhrleuSerSerClnValleasnProThrAlaIleThrValSer 932

```

```

943 ..... 943
933 AlaAlaInThrThrLeuThrAlaAlaGlyLeuThrThrProThrI 949
943 ..... 943
949 eThMetGlnProValSerGlnProThrGlnValThrLeuIleThrAlaP 966
942 ..... GTGGCGGCGCTTG 931
966 roSerGlyValAlaGlnProValHisAspLeuProValSerIleLeu 982
930 TGTAATGCGCGCTTCAATACGACCC..... 904
983 AlaSerProThrThrGlnGlnProThrAlaThrValThrIleAlaAsp 999
903 ..... GGAAATCAGCGCGCTTCTGTGCAACC 876
999 rGlyGlnGlyAspValGlnProGlyThrValThrLeuValCysSerAsp 1016
875 AATTCGCGCGCGCTTCAATACGACCCAAACGCTACGCA 826
1016 roProCysGlnThrHis..... GlnThrGlyThrThr 1026
825 GAGCGCGGCTTGTGACTTGAGAACCACTAGGCAATAC..... GC 782
1027 AsnThrAlaThrThrThrValAlaHisLeuGlyGlnHisProGlnPr 1043
781 GCTCGGCTTCAAGCGCGCTTGCACAAACA...ACGGCG..... 745
1043 oThrGlnValGlnPheValCysAspArgGlnGlnThrAlaAlaSerLeu 1060
744 ..... AATGTAATTCATCTGTAATTCAT 718
1060 aThrSerAlaValGlyGlnGlnAsnGlyAsnValAlaValArgValCys 1076
717 GGTCCACAGCGCTTATTCGCGCGCGCGCTGTAAGTAATGTCGCG 668
1077 AsnProPro..... CysGlnThrHisGlnThrGlyThrAsnThrAl 1091
667 TCCCACTCAACCGCGAGATGCGG..... GCCCGCAATTCGA 630
1091 aThrThrAlaThrSerAsnMetAlaGlyGlnHisGlySerSerAsnPr 1108
629 TGTGT..... TTGATGTTGGCAGCATTTTCAACGCGACGCTGCGCC 586
1108 roCysGlnThrHisGlnThrGlyThrThr.SerThrAlaThrThrAlaMe 1124
585 AGCTGCTTACAAACATGATTTGCGTGGTCAACGCGCTCAATTCGA 536
1124 tSerSerMetGlyThrGlyGlnGlnAlaGAspThrArgArg...ThrThrA 1140
535 ACAGGCGCG..... CGTTGAATCTCGCGCGCTTCTGTGTAATG 495
1140 snThrProThrValAlaArgIleThrValAlaProGlyAlaLeuGlnArg 1156
494 ACCGTAAGGCTGCGCA..... GCCAGCGGATTTGCT 466
1157 ValGlnGlyThrValLysProGlnCysGlnThrGlnGlnThrAsnMet 1173
465 GTCCATCGCATGACGAAGATGGCAGACGCGCTGCGCATGACGCGAGAA 416
1173 rThrThrThrMetThrValGlnAlaThrGlyAlaProCysSerAlaGlyP 1190
415 TTTGCTGACGAGCGGCTG..... CGCAGCGCATGTCACAAACGCGAT 372
1190 roLeuLeuAlaArgProSerValAlaLeuGlnSerGlySerHisSerPro 1206
371 TGGATCAGGTTGGCGGCGCACTTCTGCGCGCTTAAGTTGCGAGCGCTTC 322
1207 PheValGlnLeuAlaLeuProSerValArgValGlyLeuSerGlyProSe 1223
321 AGGTGGCTAGCGTTCAAACTCGATTTCGTGCTTCGCTTCAACGCAATCA 272

```

```

1223 rSer..... LysAspMetProThrAlaArgGlnProGlnThrThrI 1238
271 CGACTGACTGACGAGCGCGCTTTCGCCACGCGTGAATCGCGGCGATTTG 222
1238 hTrThrThrThrAsnThrProThrThrThrArgSerIle...MetValAla 1253
221 CCTGAACCGCGCGCAGTAACACACGCGCGATTTCTTTGCTTCGCAAA 172
1254 GlyGlnLeuGlyAlaAlaArgValAlaPro.....Th 1264
171 CAGCAGCTTGCGCTTTTTCAGCGCA...TCGCGCTTCTGACTTTCATCG 125
1264 rSerThrThrGlyGlnSerLeuGlnAlaSerSerProSerThrMetThr 1281
124 AGGCGCGCATACCGGATATTCCTGCGCACG..... 93
1281 eThrAlaLeuGlnAlaLeuLeuCysProSerAlaThrValThrGlnVal 1297
92 ..... AACGC 88
1298 CysSerAsnProProCysGlnThrHisGlnThrGlyThrThrAsnThrAl 1314
87 GACTTCGTAATGCGCGCGCGCTTCAACGCGCTTCTGCGGATTCGCC 39
1314 aThrThrSerAsnAlaGlySerAlaGlnArgValCysSerAsnProPro 1330

```

```

seq.name: sp_trodent:Q9QWH2
seq_documentation_block:
ID Q9QWH2 PRELIMINARY; PRT; 2045 AA.
AC Q9QWH2
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DI 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE C1 TRANSCRIPTION FACTOR.
GN HCFCL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kristie T.M.;
RT *CDNAs encoding the mouse homolog of the human transcription factor C1
(RCF).";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U08021; AAD09225.1; -.
DR MGD; MGI:105942; Hcfcl.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF01344; Kelch; 4.
DR SMART; SM00060; FN3; 1.
SQ SEQUENCE 2045 AA; 210546 MW; 7AD38FCD78EAB9C CRC64;

```

```

alignment_scores:
Quality: 109.50 Length: 617
Ratio: 0.423 Gaps: 25
Percent Similarity: 41.977 Percent Identity: 19.935

```

alignment_block:

us-09-303-518d-125/rev x Q9QWH2 ..

Align seg 1/1 to: Q9QWH2 from: 1 to: 2045

```

1343 CAGCTTCCTTCATGATGTTTCACACTTTCGCAACAGCGGCGCTA 1294
:::||||::: |||
747 LysProThrThrLeuGlyIleSerSerVal.....SerProSe 759
1293 TTGCTATTGCGCGCGCAAGCGTGCACAAACGAGGCTTCTTGTGT 1244
::: ||||| ||| ::| |||::: |||

```

```

759 rthrthlyrsproglythrthrllellelysthrilleproktsra 776
1243 ccaattccaaagaccatgctgcgcgtcgatccgcacgatt 1194
776 lallelethrlnalaglyalathrlnghlyvalthrserserproglyle 792
1193 aaatcg..... 1188
793 lyserserproillethrillethrthlyrsvalmethrserythrhl 809
1187 .....cgaaaagacgggtgcgcagca 1166,
809 yalaproalalysillethrlnalavalprolystlealathrlnghls 826
1165 tatccaaagggacacacggc..... 1146
826 lyglnghlyvalthrlnghlyvalleulysglyalaproglylnpro 842
1145 .....tcgtaagtaacaaacggcaccacggcggtgcgcggc 1108
843 glythrilleulargthrvalprometglylvalargleulvalthrpr 859
1107 gttgcac.....cgtcgttgaaacttgaaagattgt 1076
859 ovalthrvalseralavalysproalavalthrthleu..... 872,
1075 ttttcagaaatggcgcgggttcacg..... 1047
873 .....valvallyseglythrthrglyvalthrthr 882
1046 .....gtgatgagattatgcccgcgtgcgcgc 1018
883 leuglythrvalthrghlythrvalserthrsleulalaglyalaglyal 899
1017 aaccacccgaacagctcttgcgcgccttcttgatrac. ggaatc 969
899 ahiserthrsalaserleulathrproillethrthleuglythr 916
968 tgattggtgagctgcccgaatc..... 943
916 lealathrleuserseroleinvalileasnprothralallethrvalser 932
943 ..... 943
933 alaialaglnthrthleuthralalaglyleuthrthrprothrl 949
943 ..... 943
949 ethrmetglnprovalserglnprothrglnvalthrleullethrlnar 966
942 .....gtcgcgcgccttg 931
966 roserglyvalglualaglnprovalhlsaspleuprovalserlleleu 982
930 tgfatacggcggcgttcaataccgac..... 904
983 alaserprothrthrglnprothralathrvalthrillealaspse 999
903 .....ggaatcaccgcttgcctgcctcaacc 876
999 rgllynglnspvalglnproglythrvalthrleulvalcysersnpr 1016
875 aattgcggcgcagtaattgcatcttgcaccccaaacggatgcgca 826
1016 roprocysgluthrhls.....gluthrglythr 1026
825 gagggcggttgccttgacttgaaacacactgagggcaatcac.....gc 782
1027 aenthralathrthrvalalalasnleuglyghlspglnpr 1043
781 gctgcggttcagacgacctgttcgcaacaa.....acggcc..... 745
1043 othrghlnvalglnphevalcysasparglnghlthrlnalalaserleu 1060

```

```

744 .....aatgtaattacattctgattatgat 718
1060 althrseralavalglnglnghlnghlyasnvalalargvalcysser 1076
717 ggtccacacgggtttatccgcccgcgcgctcgatgaatgaatgctgc 668
1077 asnpropro.....cysgluthrhlsghlthrghlythrthrasnthral 1091
667 tgcctcaaacggcagagatgccc.....gccgcgcaatgca 630
1091 athrthralathrthrsasmetalaglylnghlsghlyserersnpr 1108
629 tctgt.....ttgatgttgacacattttcagacggcagctgcgc 586
1108 rocysgluthrhlsghlthrghlythrthr..serthralathrthralame 1124
585 agctgccttacaacacatgattttgcgttcgctcaaacgctcaatgca 536
1124 tsersermetglythrghlynglnghlnghlyasprthrarg...thrtha 1140
535 acagggcg.....cgttgaaatccctgcgcgcttcttgatgatg 495
1140 snthrprothrvalalarglilethrlnalalaproglyalaleuglnarg 1156
494 accgtagggctgcga.....gccacggatttgt 466
1157 valghlghlythrvallysproghlncysghlthrghlghlthrasmetth 1173
465 gtcacatcgacattgacagatgacgacgctgcgacatcgacgcgacgaa 416
1173 rthrthrthmetthrvalghlnalathrghlyalarglyalarglyalargly 1190
415 ttttgcgtgaacggcgggtg.....cgacggcagctccacaaacggat 372
1190 roleulathrproservalalaleuglnserghlyserthrserserprola 1206
371 tggatcaggttgcggcgacacttctgcgcgctgaattgacgagctc 322
1207 phevalghlnleulaleuoproservalargvalghlyleuiscrlprose 1223
321 agtgcgtgacggttcacaaactcgatttcgcttccttcacggcaatca 272
1223 rser.....lysasprmetprothralarglnproghlthrthrlst 1238
271 cgaactgaagtagcgcgttctccacacgctgaatccggcgatttg 222
1238 hrtyrthrthrasnthrprothrthrthargserlle...metvalala 1253
221 cctgacggcggcgcagtaaacacacacgacggcgatcttcttgcctcaaa 172
1254 glyluleuglyalalalargvalalpro.....th 1264
171 cagcacttgccctttttgacggca.....tcgcttcttgactttgatcg 125
1264 rserthrglyluserleuglnalaserproserthrsleuthm 1281
124 agggcgcatcggcagatattcttcgacac..... 93
1281 ethrlnalaleuglnalaleuucysproseralathrvalthrghlnval 1297
92 .....aacgc 88
1298 cysserasnprprocyglnthrlnghluthrghlythrthrasnthral 1314
87 gacttcggtatggccggcgccgctgaacggctgcgtgcgctgcgc 39
1314 athrthrsasnalaglyseralalargvalcysersnprpro 1330

```

seq_name: sp_gvirus:098mw2

seq_documentation_block: ID 098mw2 PRELIMINARY; PRT: 1504 AA.

AC 098MW2;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE GAG POLYPROTEIN [CONTAINS: CORE PROTEIN(S) P24].
 GN GAG.
 OS Avian leukosis virus.
 OC Viruses; Retrovirdae; Retroviridae; Alpharetrovirus.
 ON NCBI_TaxID=11864;
 RX MEDLINE=21165271; PubMed=11264350;
 RA Johnson J.A., Heneline W.;
 RT "Characterization of endogenous avian leukosis viruses in chicken embryonic fibroblast substrates used in production of measles and mumps vaccines."
 RL J. Virol. 75:3605-3612(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ENDGENOUS AVIAN LEUKOSIS VIRUS 1, EV-1;
 RA Johnson J.A., Heneline W.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO GAG GENE PROTEIN P24 (CORE NUCLEOCAPSID PROTEIN).
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO
 CC KNOWN AS THE RETROPEPSIN FAMILY.
 DR EMBL: AY013303; AAK13200.1;
 DR HSSP: P03322; 2RSP.
 DR InterPro: IPR001969; Asp.-protease.
 DR InterPro: IPR001995; Asp.-protease.
 DR InterPro: IPR000721; Gag-P24.
 DR InterPro: IPR003308; Integrase_Zn.
 DR InterPro: IPR004028; Retro_M.
 DR InterPro: IPR002156; RNaseH.
 DR InterPro: IPR000477; RYase.
 DR InterPro: IPR001878; Znf.CCHC.
 DR Pfam: PF00607; gag_p24_2.
 DR Pfam: PF02022; Integrase_Zn; 1.
 DR Pfam: PF02813; Retro_M; 1.
 DR Pfam: PF00075; RNaseH; 1.
 DR Pfam: PF00665; rve; 1.
 DR Pfam: PF00077; rvp; 1.
 DR Pfam: PF00078; rvt; 1.
 DR Pfam: PF00098; zf-CCHC; 2.
 DR SMART: SM00343; ZNF_C2HC; 2.
 DR PROSITE: PS00141; ASP_PROTEASE; 1.
 DR PROSITE: PS01175; ASP_PROT_RETROV; 1.
 KW Aspartyl protease; Core protein; Hydrolyase; Polypeptidein;
 KW RNA-directed DNA polymerase; Zinc-finger
 FT CHAIN 1 175
 FT CHAIN 178 239
 FT CHAIN 240 479
 FT CHAIN 489 577
 FT CHAIN 578 701
 FT CHAIN 709 1504
 FT CHAIN 709 1280
 FT CHAIN 1281 1504
 FT CHAIN 1504 AA; 163199 MM; 979779BA89620435 CRC64;
 SEQUENCE

alignment_scores:
 Quality: 107.50 Length: 486
 Ratio: 0.500 Gaps: 32
 Percent Similarity: 44.239 Percent Identity: 22.634

alignment block:
 US-09-303-518d-125 x 098MW2

Align seg 1/1 to: 098MW2 from: 1 to: 1504

9 AATCAAAAAGTCTTAACCTGCGCCATCGGGGAGACCGGAGCAAGCGG 58

492 AsnArgLuarArgasp.....GlyGlnThrGlySerIyol 503
 59 TTTACGAGCG.....CCGGCCATTACCGAA 84
 503 YArgAlaArgGlyLeuCySTyTrhCySgLySerProGlyHISLyGlnA 520
 85 GTCCGTTGCTGGCGCAAGAAATATGCCGGATGCGCCCTCGATGAAGT 134
 520 La..... 520
 135 CAGGAAGCGGATGCCGTCAAAAAGCCCAAGTCTGTTGAACACAAA 184
 521GlnCyProLySLySArgLySerGlyASerArgLyu.. 533
 185 AGATCCGGCGGCGTGTACTGCGCCCGCTTCAGGCAAAATCCCGCG 234
 534ArgCySgLnLeuCySaspGlyMetGlyHISAsn..... 544
 235 ATTACCGTGGCGCAAGCGCGTACTTCACTGATGATGATGCGGTTGA 284
 545AlaLySgLnCySArg... 549
 285 AGGCAAGCGCAAGATGAGTTGAACGCTACGACCTGAAGCGCTGGCA 334
 550ArgArgasp.....Gly.A 554
 335 ACTTAAGCGCGCAAGAGTGGCGCGCAACCTGATCCATCCGGTTGG 384
 554 sn...GlnGlyGlnArgProGlyArgGlyLeu...SerSerGlySerTyr 568
 385 ACTGCGCTGCGCAACCGCTGTCAGCAAAATCTCGCTGCGATGCGCA 434
 569ProValSerGlnGlnProAlaValSer..... 577
 435 GCCGTTGCGCATCTGTCATGATGATGATGATGATGATGATGATGAT 484
 578 ...LeuAlaMetThrMetGlnHISLySasp...ArgProLeuValArg. 591
 485 ACCCTAGCGGTCTTATCAAGAGAGCGCGGAGATTCAAGCGGCGCTG 534
 592ValIleLeuThrAsnThrGlySerHisProValLySgLnArgSer 606
 535 TTGCTATGAGCGCTTGTACCGCAAGCAAAATCCATGTTTAAGCAGC 584
 607 ValTyrIleThrAlaLeuLeuasp.....Se 615
 585 TGCGCGACAGCGT.....C 598
 615 rGlyAlaAspIleThrIleLeuSerGlnGluAspTyrProThrAspTyrP 632
 599 CGTGTGAATATGCTGCAACATGCAACATGATGATGATGATGATGAT 648
 632 roValMetGlnAlaAlaAsnProGlnIleHisGlyIleGlyGlyIle 648
 649 CCGTCCGTTGAGTGGCGACGACATTCATTTCATGAGCGCGTGGCGC 698
 649 PromethrArgLySer..... 653
 699 GAATAAACCGTGTGACATCATTAATATCAAGATGATTAATACCATGGC 748
 654ArgAspMetIleGlnValGly 661
 749 GTTGTGTTGACAGCGCGTGTGACAGCGCGTGTGTTGTTGTTGTTG 798
 661 alleAsnArgAspGlySerLeuGlnArgProLeuLeuLeuPheProAla 677
 799 GGTTCACAGTCAACAA.....CCGCGCTTGTGCTGCTGTTGG 842
 678 ValAlaMetValArgArgGlyTyrProArg..... 687
 843 TGGCAAGATATCCCAATTAATGCGGCGCAATGCTTGAACAGCAAC 892

456 pGlyArgPheLeuValAlaValLysGlyAlaProasp..... 468
533 TGTGTATTGAGCGCTTGACGAGCAAGCAAAATCCATGTTTAAGCA 582
469GlnLeuLeuLysArgCysLeuLeuArgAspLysAla 480
583 GCTGGCCAGACGTCGCCGTGAAAATGCGCCAC...ATGCAACACA 629
481 GlyAspLeuAlaProLeuAspGlyLysValThrAsnLeuIleArgThrAs 497
630 TGAATTGGGGGGCCGATCTGCC.....GGTTGAGTGGCAGGCACA 673
497 nasSerGluMetAlaIleGlnIleAlaLeuArgValLeuAlaGlyAlaTyr. 513
674 TTGATTTTCATGAGCGCGTGGCGGCAATAAACGCTGGACCATCAT 723
514 ..LysIleIleAspSerIleProGluAsnLeuThrSerGluLeuGlu 529
724 TATCAAGATGTAATTACCATTTGGCGCTTTGTTGCACAGCGCTGTGA 773
530 ...AsnAspLeuIlePheThrGlyLeuIle.....GlyMetIleAs 542
774 CACCGAGCCGATGATGCCCTAGTGCTTCACATCAACAACCGCGCC 823
542 pProGluArgProGluAlaIleAlaGlyAlaValArgValAlaLysGlyAlaG 559
824 TCTTGGCTACCGCTTTGGGTGGCAAGATATCGCAATTACTGCGGCGCA 873
559 LylLeuArgProIleMetIleThrGlyAspHisGlnAspThrAlaGlnAla 575
874 TTGGTT.....GACACAGCAACCGCGCT 896
576 IleAlaLysArgLeuGlyIleIleAspAlaAsnAspThrGluGlnHisVa 592
897 GATTTCCGGTTGGTATTGAACGCGCGCATTAACAAGCGCGCAGATT 946
592 LeuThrGlyAlaGluLeuAsnGluLeuSerAspGluGluPheGluLysV 609
947 ATTGGAGCGTACCAACAATCAGATTTCCGTTATGAAAGAGCGCGAGC 996
609 alValGlyIleThrSerValTyrAlaArgValSerProGluHisLysVal 625
997 AAGAGCTGTCGCGCTGG 1014
626 ArgIleValLysAlaIlePro 631
seq_name: sp_human:095769
seq_documentation_block:
ID 095769 PRELIMINARY; PRT; 915 AA.
AC 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE WIGSC:H_DJ412A9.3 PROTEIN.
GN WIGSC:H_DJ412A9.3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=9063792; PubMed=9847074;
RA Sulston J.E., Waterston R.;
RT "Toward a complete human genome sequence.";
RL genome Res. 8:1097-1108(1998).
RP SEQUENCE FROM N.A.
RA Murray J., Lennox S., Harmon G.;
RT "The sequence of Homo sapiens PAC clone RP3-412A9.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.

RA Waterston R.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC005005; AAD15619.1;
DR HSSP: 001082; 1BRK.
DR InterPro: IPR001715; Calponin_hom.
DR Pfam: PF00307; CH; 1.
DR SMART: SM00033; CH; 1.
DR PROSITE: PS00021; CH; 1.
SQ SEQUENCE 915 AA; 98919 MW; FFOAC36BDEC9B30 CRC64;
alignment_scores:
Quality: 107.00 Length: 500
Ratio: 0.46 Gaps: 31
Percent Similarity: 48.000 Percent Identity: 24.200
alignment_block:
US-09-303-518D-125 x 095769 ..
Align seg 1/1 to: 095769 from: 1 to: 915
26 ACCTGCCATCGCGGAGACCGCAAGCCGTTAGACGCCCGCC 75
111 CG.....GATGCGCCCGCCGCA 127
231 roGlySerProGluProProSerProProLysThrThrSerProGlu 247
198 ThrSerSerProAlaSerProSerSerSerProThrProAlaSerPr 214
76 ATTACG.....AGTCGCTTGTGGCAAGATATGCG 110
214 obLupProLeuGluProAlaGluAlaGlnCysLeuThrAlaIleValP 231
111 CG.....GATGCGCCCGCCGCA 127
231 roGlySerProGluProProSerProProLysThrThrSerProGlu 247
128 TGAAGTCAAGAGAGCGATCCGCTCAAAAAGCCAGCAAGTGTGTTGAA 177
248 ProGlnGluSerProThrLeuProSerThrGluGlnValValAsnLys 264
178 G..... 178
264 SLeuLeuSerGlyProLysGluThrProAlaAlaGlnSerProThrArg 281
179ACAAAAAGANTC..... 190
281 LProSerAspThrLysArgAlaAspValAlaGlyProAlaGlyProCysGln 297
191 CGGCGTGTGTTACTG.....CGCGCTTCAGGCAAAATCGCGCG 234
298 ArgSerLeuSerValLeuSerProArgGlnProAlaGlnAsnArgGlnS 314
235 ATTACCGTGGCGGAAAGCGGACTTCAGTCAAGTCAAGTTCGCTTGA 284
314 erThrProLeuAlaSerGlyProSerSerThrGlnArgAlaGlySerVal 330
285 AGCCAGCAGCAAAATCGATTGAAGCTAGCGACCTGGAAGCGGCGCA 334
331 ArgAspArgValHisLys..PheThrSerAspSerPrometAlaAlaArg. 346
335 ACTTAACGGCGGAAGAAGTGGCGGCAACCTGATTCACATCCGCTTG... 381
347LeuGlnAspGlyThrPro 352
382 TGAAGTGGCTGGGCAACCGCTTCAGCAAAATTCCTG...CCGTGA 428
353 GlnAlaAlaLeuSerProLeuThrProAlaArgLeuGlnGlyProSerL 369
429 TGGCGAGCGCTTGGCATCTGTCAATGCAATGCAACCAATCGCTGG 478
369 euthSerThrThrProAlaSerSerSerSerGlySerSerSerArgG 385
479 C...TCCGACCTACGATTCATTAACAAGAGCGCGAGGATTTCAA 525
385 yProSerAspThrSerSerArgPheSerLysGluGlnArgGlyValAlaG 402


```

526 CGGGGCTTGGTATGACCGGTTGACCGAAGCAATCATGTTTG 575
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
402 InProleuAlaInleuArGserGProGInGInGPro.ArGgl 418
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
576 TAAGCAGCTGGCGAGAGCTGCTGAAAATGCTCCAC... 618
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
418 yArGgIyLeuAlaIaArG...ProleuAlaInArGAlaIyGlyProV 434
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
619 ...ATCGAAGACATGAATTCGGCGCGCATCTCGCGTTGAGTGGC 666
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
434 alAlaArGserGInGInGProGAlaIaProleuProValAlaVal...Gly 449
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
667 ACCGACATTCATTCATGACCGCGCGCGAATAAACCGTGTGAC 716
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
450 Thr.....AlaGInProGlySerMetLysThrThrPheTh 462
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
717 CATCATTTATCAAGATGTAATTAACATTCGGCGTTGTTGACAGCGC 766
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
462 rTlGgIuIleYsAsP.....GlyArG.....GlyG 471
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
767 GTCGTAACACGAGCGCGTATTCGCGTGTGTTCACTCAACAA 816
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
471 InAlaSerThrGlyArGValLeuLeuProThrGlyAmGInArGAlaGlu 487
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
817 .....CCGCGCGCTTGGCGTACCGGTTTGGG 842
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
488 LeuThrLeuGlyLeuArGAlaProProThrLeuLeuSerThrSerG 504
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
843 TGGGAAGTATCGCAATTCATTCGCGCGAATTCGTTGACAGACAC 892
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
504 yGlyLys...SerThrIleThrArGValAlaInSerProGlyThrLeuAla 520
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
893 GCGTATTCGCGTTCGTTATTCAGCGCGCATTCACAGCGCGCAC 942
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
520 rGleu.....GlySerValThrHis.....ValThrSerPheSerHis 532
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
943 .....GATTAATTGGAGCGCTACCAATCAGATTT.....C 974
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
533 AlaProProSerSerArGly.GlyCysSerIleLysMetGInAlaGlu 549
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
975 CGTTATCGAGAGAGCGCGACAAAGAGCTGTCGGCTGGGTTGCGCGC 1024
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
549 rAlaGInProleuAlaIaAlaValGInAlaAlaInGlyAlaGInGIn 565
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1025 AGCGCGACAAT...ACTCATCAGCGGTACACCGCTCGGCGCATTCG 1071
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
566 ThrArGValAlaInLysAlaPro.GInGlyArGserProleuSerAlaGlu 582
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1072 AAAAAACAATCTTCAA..... 1088
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
582 InLeuMetThrIleGluAspGluGlyValLeuAspLysMetLeuAspGln 598
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1089 .....GTTCAGACAG 1099
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
599 SerThrAspPheGInGluArGValLysLeuIleArGAlaAlaLeuArGVal 615
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1100 CCGTCAACGGGCGACCGCGCATTCGCGATTCGTTACTTACGAGCGC 1149
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
615 uArGgInArGysArGAspGInArGAspLysGInArGAlaArGLeuG 632
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1150 GTGATCGCTGGATATCCGCGACCGCTGCT.....TTTGGCGCATTT 1193
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
632 InGluAlaArGgIyArGProGInGlyGInGlyAsnThrAlaThrGlu 648
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1194 AATCGTCGCGATTCAGAGCGCGAGCGAGCGCATTCGGG 1229
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
649 ThrThrThrArGHisSerGInArGAlaAlaAspLys 660
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
seq_name: sp_invertebrate:076602
seq_documentation_block:
ID 076602 PRELIMINARY; PRT; 1275 AA.
AC 076602;

```

```

DT 01-NOV-1998 (TREMUREL. 08, Created)
DT 01-NOV-1998 (TREMUREL. 08, Last sequence update)
DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)
DE HYPOTHETICAL 122.9 KDA PROTEIN.
GN H02F09.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Anisconough R., Anderson K., Baynes C., Berts M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green F., Hawkins T., Laister N., Latreille P.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMuray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Koopra A., Saunders D., Showken R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Geisel C., Harmon G.;
RT "The sequence of C. elegans cosmid H02F09."
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077538; AAC64622.1; -.
KW Hypothetical protein.
SQ
SEQUENCE 1275 AA; 122924 MW; CB59B70C05959E25 CRC64;

alignment_scores:
Quality: 107.00 Length: 431
Ratio: 0.500 Gaps: 16
Percent Similarity: 49.652 Percent Identity: 21.578

alignment_block:
US-09-303-518D-125/rev x 076602 ..
Align seg 1/1 to: 076602 from: 1 to: 1275
1319 AGCACTTGGCGACAGAGCGCGATTCGTTATTCGCCGCGACAGCA 1270
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
218 AsnAsnIleLeuAsnIleGlyThrThrGInThrProProValThrThrSe 234
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1269 GCTGCACAAAGCGAGTCTCTCGTCCATTCACAGCAACCAAT...G 1223
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
234 rThrMetaIaThrThrThrAlaAsnValThrSerAlaIaProAsThrT 251
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1222 CCTGCGCGCTGCGTATTCGCGACGATTAATCGCGCAAAAGCAGGTG 1173
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
251 hrValThrIleSerThrSerProThrThrValThrValProSerThr 267
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1172 GCGAGATATTCAGAGGATCAGCGCGCTGCTAGTACCATTCGCGACAT 1123
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
268 AlaGInThrSerSerThrThrValThr...ValProThrThrThrVa 283
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1122 GCGCGGTCGCCCGCGGTG.....ACGGCTGTGTGAACCTGA 1085
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
283 ThrGlyProThrThrValValThrValProThrThrValThrValThr]e 300
    :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1084 AGAGTTTGTTCAGGAATGCGCGAGGTTGACGCGTGAATGAGATAT 1035

```

```

300 roSerThrValValThrSerProIleThrThrProSerThrValValThr 316
      |||
1034 TTGTCGGGCTCCGGCGACACCGCAACAGCTTTGCTGGCGCTTC 985
      |||
317 ValProSerThrValValThrValProSerThrAlaValThrLysProse 333
      |||
984 TTCGATACGGAATCTGATTGTGGTCCCAATATCGTGGCCG 935*
      |||
333 ThrValValValValValValValValValValValValValVal 339
      |||
934 CTGTGTAAATCGCGCTCAATACCGAACCGGAATCAAGCGGTGTCT 885
      |||
339 roSerThrValValThrValProSerThrValValThrLysProSerThr 355
      |||
884 GTGTCAACCAATCGCGCGAGTAATTCGATACCTTTCGCCACCAAAAC 835
      |||
356 ValValThrSerSerProThrValValValValValValValValVal 369
      |||
834 GTACCGACAGGCGCGGTTGTGACTTGAGAACCCCTAGGCAATCA 785
      |||
369 ValValValThrThrProSerThrValValValValProSerThrVal 386
      |||
784 CGGCTCGGTGTACAGCGCGCTTGCAAAACAGCCCAATGTAAAT 735
      |||
386 hrValValValValValValValValValValValValValValVal 399
      |||
734 ACATCTGATTAATGATGTCACAGCGCTTTATTCGCGCGACCGGCTC 685
      |||
400 ThrAlaValValValValValValValValValValValValValVal 412
      |||
684 GATGAATGATATGTGCTGCCACTCAAAACCGGAGATGCGCGCGCA 635
      |||
412 lMetValValValValValValValValValValValValValVal 414
      |||
634 ATTCATGTGTTCGATGTGGCAGCATTTTCAGACGCGACGTCGGCGCA 585
      |||
414 hrSerArgSerThrValIleThrThrProThrThrIleGlySerSerPro 430
      |||
584 GGTGCTTACAAACATGATTTGCGTTCGCTCAAAACGCGCTCAATACCA 535
      |||
431 SerThrAlaGlyThrValValValValValValValValValValVal 435
      |||
534 CAGCGCGGCTTGAATCTCGCGGCTTCTTGTATGATACCGCTAGGCT 485
      |||
436 SerLeuAlaSerThrAlaValThrThrGluThrSerIleGlyS 450
      |||
484 CGGACGCGACGCGATG.....GTGTCCATCGCATGACGAG 447
      |||
450 erSerSerThrProLeuProSerGlnSerThrSerLeuSerMetSerSer 466
      |||
446 ATGCGG.....AAGGCTCGCATCGACGCGAGAGATTTGCTGAACG 403
      |||
467 LeuSerThrThrProSerSerSerThrAlaIleValThrSerProAl 483
      |||
402 ACGGCTCGCGACGCGACGCTCAAAACCGGATGATGAGCTTGGCGCA 353
      |||
483 aThrGlnGlnSerThr.....LysProThr.....IleGlyThrSerMetS 497
      |||
352 CTCTGTGCGCGCTTAAGTTGCCAGCGCTTCAGGTGCGCTCAAC 303
      |||
497 erSerGlyProThrThrValAlaProIleAlaSerThrGlnSerThrVal 513
      |||
302 TCGATTTCGTGCTGCTTCA.....ACGGCAATACGACGTACGAGTAC 256
      |||
514 LeuGlnSerSerThrProSerGlyThrThrValThrLeuProSerLysE 530
      |||
255 GCGGTTTTCGCCACGCGGATTCGCGGATTTTTCCTGAAGCGCGAG 206
      |||
530 rSerThrAla.....ThrAlaGlyThrSerProGlnAlaSerThrHy 544
      |||
205 TAAACACACGCGCGGATTCCTTTCTTCAACAGCACTTGCGCTTTT 156
      |||

```

```

544 alThrThrValThrAspIleSerThrValSerLysSerThrValThrSer 560
      |||
155 TTGACGGCAGTCGCTTCTTTCATTCATGAGGCGCGATACCGGATTA 106
      |||
561 GlnThrAlaGlnSerSerLeuSer.....ThrGlnSerProThrSerAla 576
      |||
105 TTCCTCCCAAGCAACGCGACTCGGTAATGCGGCGCGCTCG 63
      |||
576 ySerSerIleSerThrValSerThrValSerGlnProSer 590
      |||

```

seq_name: sp.bacteria:005089

seq_documentation_block:

```

ID 005089 PRELIMINARY; PRT: 545 AA.
AC 005089;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHEtical 57.3 KDA PROTEIN.
OS Nocardioides simplex (Arthrobacter simplex).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Propionibacteriaceae; Nocardioidaceae;
OX NCBI_TaxID=2045;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO12069;
RX MEDLINE=95319331; PubMed=7596291;
RA Molnar I., Choi K., Yamashita M., Murooka Y.;
RT "Molecular cloning, expression in Streptomyces lividans, and analysis
RT of a gene cluster from Arthrobacter simplex encoding 3-
RT ketosteroid-DELTA-1-dehydrogenase, 3-ketosteroid-DELTA-5-isomerase
RT and a hypothetical regulatory protein."
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z93338; CAB07541.1; -
KW Hypothetical protein.
SQ SEQUENCE 545 AA; 57318 MW; 55F41EBD5E3A19E CRC64;

```

alignment_scores:

```

Quality: 109.50 Length: 493
Ratio: 0.464 Gaps: 33
Percent Similarity: 47.870 Percent Identity: 25.152

```

alignment_block:

us-09-303-518d-125 x 005089 ..

Align seg 1/1 to: 005089 from: 1 to: 545

```

18 AGGTCTAATCTGCCAT...CGCGGACAGCGGAGCAAGCGCTTAGG 64
      |||
61 ArgSerArgGlyAlaGlnValArgGlySerAlaSerArgSerIleYargAr 77
      |||
65 ACGCGCGCGCATTACGA.....AGTCGCTTGTCTTGC 99
      |||
77 gSerSerGlyArgArgArgGlyScyScyProGlyGlySerArgArgArgCySA 94
      |||
100 GAAGATATCGCGTATGCGCCCTCGATGAAGTAAGTAAGAGG...CGA 146
      |||
94 rArgArgArgIleArgValGluPro.....AlaGlnIleYargValasn 107
      |||
147 TCGCGTCAAAAAAGCGCAAGTGTCTTTGAAGCAAAAAAGATCCGGCG 196
      |||
108 AlaSerSerGlyArgProAlaAla.....ArgAlaSerSerAlaIar 121
      |||
197 TGGTGTTTACTGC...GCGGCTTCAGGCAAAATCGCGGATTCACCGT 243
      |||

```

```

121 g.....CysArgAlaGlyAlaArgSerProArgThrProG. 135
244 GG.....CGAAGCGCGT..... 257
135 LylGlyProSerAlaGlyArgAlaSerArgAlaArgSerCys 151
258 ...ACTTCAGTCAGTCGTGATTCGGTTGAAGGCAACGCAAAATTCAGT 304
152 GlyThrThValGlnAspGln.....ProHisArgAlaAspArgVal 165
305 TTGACGCTACGCACTGCAAGCGCTGGCAAAATTCAGCGGCAAGATG 354
165 L.....AlaGlyProValAlaAlaGlnHisGlyL 175
355 CGCGGCAACCTGATCCATC...CGGTTGTGACTGCGCTGCGCAACCG 401
175 euProGlnGlyHisValValGlnArgLeuValAspGlnProAlaVal 191
402 TCGGTTGAGCAAAATTCCTGC...CGTCAGTCGCAAGCGCTGGCATCT 448
192 AlaValAspHisArgAlaGlnArgAlaLeuGlnGlnHisLe 208
449 TCGTCATGCGATGACACCAATCCGCTGGCTGCGCAACCTGAGGCAT 498
208 uArgArgProAlaGlyGlnArgGlyAspGlyGlyProProGly.Leu 224
499 ATCAAGACGCGCGCGAGATTCAAACGCGCTGTTGATTCAGACCG 548
225 ValHisGln.....ArgHisSerGlyAlaHisLeuGlyL 236
549 TTTCAGCGCAACGCAAAATTCATGTTGTGAAGGCACTGCGCAACGCTG 598
236 LylLeuAspArgValAlaGlyAlaValAlaLeuValAlaGlnAlaProVal 253
599 CGTCGAAATGCTGCCACATGCAAAACATGATTCGGCGCGCGCAT 648
253 aLglnGly..... 255
649 CGTGGCGCTTGAAGTGCACG...CACATTCAT...TTTCAGACGCGGT 692
256 ...LeuGlyLeuValAlaValProHisValHisValAlaValLysProAl 271
693 CGCGCGCAATGAACCGTGGACCATCATCATATCATGATTAATACCA 742
271 aGlyArgGlnAspArgAlaArgProAlaValGlnHisGlyAspProAlaVal 288
743 TTGGCGCTTGTTCGCAACGCGCTGTAACACCGCGCGATTCGC 792
288 aLAlaLeu.....GlnHisArgAlaAspHisProValAla 299
793 CTAGGTGTTCTCAAGTCAACAAACGCGCTCTG..... 828
300 Leu...GlyAspGlnLeuAspGlnArgGlyLeuValLarProAspArgAla 315
829 .....CGTACGTTTTCGGTGCAGAAAGTATCGAAATTCATCGCG 868
315 aGlyAlaGlnArgAlaValGlnGlnProGlyArgGlnArgLeuProAlaG 332
869 GCGAATGTTGACACACAGCAACGCGGTATTCGCTGGTGGTATTCAC 918
332 LylGlnValAlaAlaAspHisHis.....AlaAlaAspAlaLeuSer 346
919 GCGCGCATTCACAAAGCGCGCATTTATTCGGACGCTACCAATCA 968
347 GlyAlaAlaGlnAspArgAlaGlnAlaLeuAlaGlyLeuAlaArgAspG 363
969 GATTC.....CGTTATGCA..... 983
363 nValHisProLeuValValArgThrGlyAspArgHisArgAspArgCysL 380
984 .....AGAAGCGCGAGCAA.....AGA 1001
380 euAspAspAlaArgProGlnGlnArgProGlyLeuAlaGlnHisArgArg 396

```

```

1002 GCTGTCCGCTGGTGGTGGCGCGGCAACAAATTCATCATCAGCGTA 1051
397 ValGlnArgLeuAlaLeuAspAlaLala..... 405
1052 CAACCGTCGCGCATTTCTCGTAAACAAATCTTCAGTT..... 1091
406 ...ProArgGlyValAlaAlaArgGlnLeuArgValValGlyAla 421
1092 .....CAACACAGCGCTACAGCGGCGGA...CGCGCGCAGTGGCC 1130
421 LaAlaArgProHisGlnLeuGlnArgArgArgProProGlnHis..... 435
1131 GATGTGATTCAGACGCGGCTGATTCCTGGA..... 1163
436 .....ProAspGlyLeuGlyAlaValProGlnGlnArg 446
1164 .TATCCTGCC.....CACCTGCTTTTTCGCGGATTTA 1194
446 gleuProAlaGlyProSerGlyArgArgHisProProProProAlaGly 463
1195 ATGCTGCGGATACGA.....CAGCGCGCAGCATTCGCTTG 1232
463 rGSerArgArgAlaArgAlaCysArgAlaSerArgAlaSerValSerAla 479
1233 ...CTTGAATTCGACGAGCAAGACCT 1256
480 GlyLeuAlaAlaSerArgArgArgPro 488
seq_name: sp_human:Q96HA1

```

```

seq_documentation_block:
ID Q96HA1 PRELIMINARY; PRT; 984 AA.
AC Q96HA1;
DT 01-DEC-2001 (TREMBLrel. 19, created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:3792).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LYMPH, AND LYMPHOMA;
RA Straussberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC008794; AAH08794.1;
SQ SEQUENCE 984 AA; 99036 MW; 36DEDB4419EBA86 CRC64;

```

```

alignment_scores:
Quality: 106.50 Length: 442
Ratio: 0.525 Gaps: 21
Percent Similarity: 45.928 Percent Identity: 21.946

```

```

alignment_block:
US-09-303-518D-125/rev x Q96HA1 ..

```

```

Align seg 1/1 to: Q96HA1 from: 1 to: 984

```

```

1313 TTGCCACACAGCGCGCTATTCGTATTCGCGCGGAGACAGAACTGCA 1264
||| :::::::::::::: ::::::::::::::
374 LeuSerGlnSerGlyProProGlyLeuLeuPro..... 384
1263 CAAGCGAGGCTTTGTCATTCACACCAACCAATGCG...TGCG 1217
||| ||| :::::::::::::: ::::::::::::::
385 .....SerProSerPheAspSerLysProProThrThrLeuLeuG 398
1216 CGCTGCGGTATCGCGCAGATTAATGCGCAAAAGCGGTGGCAGG 1167
||| ::::::::::::::
398 LylLeuLLeuProAlaProSerMet..... 405

```

```

1166 ATATCAAGGACATCACCGCTCGTAAGTACCAATCGGACCATGCGCG 1117
406 .....ValPro...AlaThrAspThrLy 412
1116 GTCCGCGCGCTTGACGGCTGTGTGAACCTGGAAGCTTTGTTTTCAGA 1067
412 salalProthrLeuGlnalaglnThrAlaThrLysProGlnalThrS 429
1066 AATGCGCGAGGGTTGTACCGCGTGGAGATTTTTCGCGCTCGCGCGCA 1017
429 eraAlaProserProAlaProLysInserPheLeuPheGlyThrGlna 445
1016 ACCGACCGCAACCGCTCTTGTGCGCGCTTCTTCATTAACGGAATCTG 967
446 ThrSerProSerSer.....ProAlaAla..... 453
966 ATTGTGTAGAGTCCCAATATGTCGCGCGCTTGTGTATGCGCGCT 917
454 .....ProAlaAlaSerSerAlaProPheMetPheLysProI 466
916 TCAATACCGCAACCGGAATCAGCGGTTGTGTGTCAACCAATTCGCC 867
466 lePheThrAlaPro..... 470
866 GCAATATTTGCGATTACTTTCGACCCCAACGCTACGCAAGCGCGG 817
471 ..... 477
816 TTTGTGACTTGTAGAACCACTTGGGCAATCAACGCGCTGCGTGTCA 767
477 yProthr.....ProGlyProSerAlaThrAlaThrAla..... 489
766 GGGCTGTGTGAACCAACGCGCAATGTATTAATCAATTTGATTAATG 717
490 ..ProSerSerSerLeuProThr..... 498
716 GTCCACACGCTTTTATTCGCGCGCAACGCGCTGATGAATGATGCG 667
499 .....ThrSerThrThrAlaProThr..... 505
666 GCCACCAACGCG.....GCAGATGCGGCGCGCAATCAATGATGCTT 623
506 .....PheGlnProValPheSerSerMetGlyProAlaSer...ValP 520
622 CGATGTGTGCGACATTTTCAGCGCGCTGCGCGCGCTTCAACAA 573
520 roleuProAlaProPhePheLysGlnThrThrProAlaThrAlaPro 536
572 ACATGATTTTTCGTTGCGTCAACGCGCTCAATACCAACGCGCGCTT 523
537 Thr..... 537
522 GAATTCGCGCGCGCTTCTTGTATTAATGACGCTGCGCGCACCG 473
538 .....ThrThrAlaProLeuPheThrGlyLeuAlaSerAlaThrSer 552
472 GATTGTGTTCATCGCATTTGACGAAGATGGGACGCGCTGCGCATCG 426
552 laValAlaProIleThrSerAlaSerProSerThrAspSerAlaSerLys 568
425 ..ACGCGAGAAATTTGCTGACGAGCGGGTGGCGACGCTCAACAA 379
569 ProAlaPheGlyPheGlyIleAsnSer...ValSerSerSerSerAlaSe 584
378 ACCGATTTGATCAGTTGCGCGCGCTTCTGCGCGCTTAACTTTGCCA 329
584 rThrThrThrSerThrAlaThrAlaAlaSerGlnProPheLeuPheGly 601
328 GC.....GCTTCAAGTGTAGCGTCAACATCGATTTGCTGCTG... 288
601 laProGlnAlaSerAlaAlaSerPheThrProAlaMetGlySerIlePhe 617
287 .....CCTTCAACGCGCAATCAGCACTGACTGAAGTACGCGCTT 250

```

```

618 GlnPheGlyLysProProAlaLeuProThrThrThrThrValThrThrPh 634
249 TTCGCGACGCGTAATCGCGCGATTTTTCCTGGAACCGCGCGAGTAACA 200
634 eSerGlnSer...LeuHisThrAlaValProThrAlaThrSerSerSer 650
199 CCAG.....CCGGAATTTTTCCTTCACAAACGACACTTGCCTTTT 156
650 laAlaAlaPheSerGlyPheGlySerThrLeuAlaThrSerAlaProAla 666
155 TTGACGCGCATCGCTTCTTCTGACTTTCATCAGAGGCGCATACCGGCA 106
667 ThrSerSerGlnProThrLeuThrPheSerAsnThrSerThrProThrPh 683
105 TTTCTCCCAAGCAACGCGCACTTCGGTAAATGCGCGCGCTGCTTAACG 56
683 eAsnIleProPheGlySerSerAlaLysSerProLeuProSerTyrProG 700
55 CTGCTCGCGTCTGCGCGCGGATGCG 30
700 LysAlaAsnProGlnProAlaPheGly 708

```

seq_name: sp_archaeo:097A73

seq_documentation_block:

```

ID 097A73 PRELIMINARY; PRT; 592 AA.
AC 097A73;
DT 01-OCT-2001 (TREMblrel, 18, created)
DT 01-OCT-2001 (TREMblrel, 18, last sequence update)
DE 01-DEC-2001 (TREMblrel, 19, last annotation update)
DE ACYLAMINOACYL-PEPTIDASE.
GN TYG0965471.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;
OC Thermoplasma.
OX NCBI_TaxID=50339;
RN [1]
RP
RC STRAIN=GS1 / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.,
RT "Archaeal adaptation to higher temperatures revealed by genomic
RT sequence of Thermoplasma volcanium."
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
DR EMBL; AP000994; BAB60079.1;
DR InterPro; IPR000379; Est_Lip_thioest_actsite.
DR InterPro; IPR001375; Peptidase_S9.
DR Pfam; PF02230; abhydrolase_2; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
KW complete proteome.
SQ SEQUENCE 592 AA; 67733 MW; B83C1FABA90FOFCE CRC64;

```

alignment_scores:

| Quality: | Ratio: | Length: |
|----------------------------|--------------------------|----------|
| 106.00 | 0.555 | 437 |
| Percent Similarity: 43.707 | Percent Identity: 21.510 | gaps: 25 |

alignment_block:

US-09-303-518D-125 x 097A73 ..

Align seg 1/1 to: 097A73 from: 1 to: 592

```

82 GAATCGCGCTTCTTGGC.....GAAGATATCGCGATGCG 119
111:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166 GlnLeuAlaPheIleGlySerArgGlyGlnGlnGlnGlnGlnGlnGln 182
120 CCCCTGATGAAGTCAAGGA.....GGCGATGCGCTCAAAAAA... 159
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

182 gProSerIleTyrValIleAspAlaGlyLysPheIleGluTyrHisA 199
160 ..... GGGCAA 165
199 rGProLysAspArgGlnLeuGlyLysLeuThrIleGlyThrSerGlyGlu 215
166 GTCGCTGTTT..... 174
216 IleAlaPheIleGluSerThrMetSerAspArgGlyValValSerGlyAs 232
175 ..... GAAGACAAAACAT..... 189
232 pValIleValLeuSerAspSerAspLysLysAsnLeuThrGluGlyAsnG 249
190 ..... CCGGCGGTGTTTACTGCGCGCTTCAGGCAAAATC 228
249 LuSerTyrSerThrIleValPhe..... TyrAspGlyGluLe 262
229 GCGGCGCAT..... CACCGTGCGGAAAGCGCGTACTTCAGTCAGTCGT 272
263 TyrValLeuGlnAsnHisGluThrHisPheArgValIleSerLeuLeuTh 279
273 GATTGCCGT... GAAGGCAAGCAGCAAAATCGAGTTTGAACGTCAGCAC 319
279 rArgLysValLeuTyrPserGlyAspGlyIleValTyrProAlaTyrSerP 296
320 CTGAA..... GCGCTGGCAAACTTAAAGCGC 345
296 roSerPheHisIleSerSerGlyLysLeuAlaLeuAla... PheSerSer 311
346 GAAGAAAGTGGCGGCAACCTGATCCAAATCCGTTGTGACGTCGTCG 395
312 ProAspGlnProGlnGluValIleValLysAsnIleAspThrGlyValAl 328
396 CACCGTCGCTTCAGCAAAATCTCGCGTCGATGCGGAGCGCTTCGCCA 445
328 uGluArgSer..... AlaI 333
446 TCTTCGTCATGATGCATGACACCAATCCGTCGTCGCGACCTAGGAC 495
333 LeuAsnSerGlyLeuLeuAsp..... LeuLysAlaTyrProSerGlu 346
496 ATTATCAAGAAAGCGCGGAGATTCAAACGCGGCGCTGTGGTATTGAG 545
347 IleValIleGluTyrLysAlaSerAspGlyLysIleTyrGlyLeuLeuAr 363
546 CCGTTTGACGCAAGCAAAATCCATGTTTGAAGCGAGCTGGCGCAGACG 595
363 gThrLeuAspProLys..... 368
596 TGCCCTCGAAATGCTGCCAATCGAATCATGATTCGGGCGCG 645
369 ..... AsnProLeuIleValTyrValHis..... GlyLysPro 379
646 ..... CATCTGCGCGGTTTGAAGTCGACGACGACATTCATTTCAT 683
380 ThrSerPheSerTyrGlyAlaPheLeuAspArgPheSerValTyr..... 394
684 CCAAGCCGTCGCGGCGGATTAACCGTCGACCATCATATATCAAGATG 733
395 ..... LeuGlyTyrGlyPheSerValPheMetProAsnIleTyrGlyLys 409
734 TAATTACCATGCGCGTTTGTTCACACAGGCGCGCTGACACCGGACGC 783
409 eTyrValGlyLeuGlyLysArgGluTyrAlaGluSerAsnIleGlyAsp..... 423
784 GTGATTGCCCTAGTGTTTCACATCAACAACCGCGCTCTTTCGCGAC 833
424 ..... LeuGlyLysMetAspPheGluAsp..... Val 432
834 CGTTTGGTGGAAGATGATCGCAATATCTGCGGCGAATGTCGTCACA 883
432 ILeSerGlyIleLysTyrLeuGln... GlnSerGlyLys... IleAspT 447

```

```

884 CAGACACCGCGCTGATTTCCGCT..... 906
447 hTyrAsnIlePheIleThrGlySerTyrGlyTyrMetSerAla 463
907 ..... TCGTATTGAAAGCGCGCATTCACAA... GG 935
464 LeuAlaValMetLysThrAspIlePheAsnAlaSerValSerLeuPheGl 480
936 CGCGCAGATTAATTTGGACGCTACCAATCAATATTCGTTATCGAAG 985
480 yLleSerAspTrpIleSer... PheHisGlyThrSerAsnLeuTyrGluT 496
986 AAGGCGCC... AGCAAGAGCTGTCGCGTGGCGCGCAGCCGAC 1032
496 rPAspArgGluIleHisLeuAspAlaAspProTyrPserPheGluLysTyrAsp 512
1033 AATATCTCATGACGCGTACAAACCTCGCCATTTCTGAAAACAAACT 1082
513 ArgTyrSerProIleArg..... IleLysAlaGlyLysP 523
1083 CTTCAGATTCAACACACAGCGCGTCAAGCGGCGCGCCATGTCGCGCA 1132
523 oLysThrProValLeuLeuMetHisGlyValAsnAspLysTyrValProI 540
1133 TTGCTACTTAC 1143
540 IleGlyGlnTyr 543

```

seq_name: sp_invertebrate:Q9BKV7

seq_documentation_block:

```

ID Q9BKV7 PRELIMINARY; PRT; 1325 AA.
AC Q9BKV7;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PPG3.
GN PPG3.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIDELIN;
RA Myler P.J., Sisk E., Cavithra J., Handley F., Vogt C., Robertson L.,
RA McDonald P., Ivens A., Nguyen D., Munden H., Stuart K., Worthey E.A.,
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AC084329; AAK31375.1; -.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003591; LRR_tyr.
DR Pfam: PF00560; LRR_6.
DR PRINTS: PR00019; LEURICHRPT.
DR SMART: SM00369; LRR_tyr; 5.
SQ SEQUENCE 1325 AA; 13385 MW; 8898928BAEF9418F CRC64;

```

alignment_scores:

```

Quality: 106.00 Length: 438
Ratio: 0.500 Gaps: 13
Percent Similarity: 48.402 Percent Identity: 20.776

```

alignment_block:

US-09-303-518d-125/rev x Q9BKV7 ..

Align seg 1/1 to: Q9BKV7 from: 1 to: 1325

```

1337 TCCCTTCATAGTGTTCAGCATTTTGGCAACAGCGCGCTATTGCTA 1288
111 111 111 111 111 111 111 111 111 111 111 111 111
741 SerSerSerSerAlaProSerAlaSerSerSerSerAlaProSerSer 757
1287 TTGGCCGCGGACAGCAAGCTGACAAAGAGTCTTCTTCGTCAT 1238
111 111 111 111 111 111 111 111 111 111 111 111

```

```

757 rSerSerAlaProSerAlaSerSerSerSerAlaProSerSerSerSers 774
1237 CCAAGCAACCAATGCCTCGCGCGTTCGGTATCGCCGACGATTAATCG 1188
774 erSerAlaProSerAlaSerSerSerSerAlaProSerSerSerSer 790
1187 CGAAAAGCAGGGTGGGAGGATATCCAGGGCAGTCACGCGCTGTAAGT 1138
791 AlaproSerAlaSerSerSerSerSerAlaProSerSerSerSerAla 807
1137 ACCAATGGCAACCATGGCGGGTCCGCCGCTTGACGCGCTGTTGTAAGT 1088
807 aProSerAlaSerSerSerSerAlaProSerSerSerSer..... 820
1087 TGAAGATGTTGTTTCAGAAATGGCCGAGGTTGTACGCGTGATGGAG 1038
821 ..... 821
1037 TATTTTCGCGCTCGCGCGCAACCGCAGACAGCTCTTTCGTGGCGG 988
822 AlaproSerAlaSerSerSerSerAlaProSerSerSerSerAlaPro 838
987 TTCTTGATACGGAATCTGATGTGTGACGCTCCCAATATATCGTCG 938
838 oSerAlaSerSer.....SerSerA 845
937 CGCCTTGTATTCGCGCGCTTCATACCGAACCGGAATCAGCGGTTG 888
845 lProSerSerSerSerSerAlaProSerAlaSerSerSerSerAlaPro 861
887 TCTGTGTACCAATTCGCCCGCAGTAATTTGCGATCTTTCGACCCCA 838
862 SerSerSerSerSerAlaProSerAla...SerSerSerSerAlaProse 877
837 AACGGTACGACAGAGCGCGGTTGTGACTTGAGAACCACTAGGGCAA 788
877 rSerSerSerSerAlaProSerAlaSerSerSerSerAlaProSerSers 894
787 TCACCGCTCGGTTCAGACGGCGCTGTCMAAACCAACGCGCAATGTA 738
894 erSerSerSerAla.....ProSerAlaSerSerSerSerAlaPro 907
737 ATTACATCTTGATATTGATGTCCACACGTTTATTCGCCGCGACGG 688
908 SerSerSer.....SerSerSerAlaProSerAl 917
687 CTCGATGAATGAATGTCGCTGCACCTCAACCGCAGAGTGGCGGCGC 638
917 aSerSerSer.....SerAlaProSerSerSerSerSerSerAlaPro. 931
637 CGAATTCATGTGTTGCATGTTGGCAGCATTTTCAGACGGCAGCTGTGG 588
932 .....SerAlaSerSerSerSerAlaProSerSerSerSerAla 946
587 CCAAGCTCCTTACAAACATGATTTTGCCTGCGTCAACGCGCTCAATAG 538
947 ProSerAlaSerSerSer.....SerAlaProSerSerSerSe 959
537 CAACAGCCGCGTTTGAATCCCTCGCGCGCTTCCTTGATATGACCGTAG 488
959 rSerAlaProSerAlaSerSerSerSerAlaPro..... 970
487 GGTGGCAGCGCAGGATGTGTCATCGCATGTGACGAAAGATGGCGAAC 438
971 ..SerSerSerSerSerAlaProSerAlaSerSerSerSerAlaProSer 986
437 GGTGGCATCGACGCGAGGAATTTTGTCTGAACGAGCGGTGCGCAGCGC 388
987 SerSerSerSerSerAla..... 992
387 AGTCACAAACCGGATTTGATCAGGTGCGGCGCACTTCTTCGCGGCTTA 338
993 .....ProSerAlaSerSerSerSerAlaProSerSerSerSers 1006

```

